秩序創発特性を有する問題解決システムの構築に関する研究

1998年

堀内　匡

論文要旨

我々の対象とする問題や取り巻く環境の複雑・大規模化に伴い，従来の中央集権型システムの限界が指摘されており，自律分散型システムあるいは創発的システムの設計に注目が集まっている．このようなシステムでは，従来のトップダウン的なシステム設計ではなく，秩序形成や創発性などの秩序創発性を内包したボトムアップ的なアプローチをより積極的に活用したシステムの構築が求められている．すなわち，生物システムのように，構成要素の協調によって大域的な秩序として機能／構造が生み出され，それが環境の変化とともに更新される特性である“秩序形成”と“創発”といった秩序創発のメカニズムを備えたシステムの構築方法の確立が強く望まれている．

本研究においては，基礎概念としてこのような秩序創発を取り上げるとともに，秩序創発を自律的秩序形成と創発的秩序形成の二つに大きく分けて考え，それぞれの原理に基づくシステムによる「問題解決」に関する考察を行なう．すなわち，前者に基づくシステムを「システム内部における秩序形成による問題解決」のためのシステムとして，後者に基づくシステムを「システム－環境間における秩序形成による問題解決」のためのシステムとして位置づけて議論する．

自律的秩序形成による問題解決システムとしては，記号処理に基づく上位層と自律分散型処理を行う下位層の二階層から構成される制約指向型自律分散システムを提案し，連続変量・ファジィ制約を含む制約充足問題への適用を行ない，機械設計問題の解決を試みる．さらに，これらのアプローチの根底にある，ファジィネスを介した連続変量の記号化と逐次的に制約を辿る制約伝播に着目し，制約指向問題解決プロセスに潜む複雑性について明らかにする．

創発的秩序形成よる問題解決システムとしては，適応学習の機構を組み込んだ創発型システムを考え，ファジィ推論を導入することにより連続値入出力を扱う強化学習について提案し，制御問題への適用を通してその有効性を確認する．また，学習の効率化・高速化を図るために経験強化を考慮した強化学習について提案し，自律移動ロボットを含むいくつかの制御問題への適用を通してその有効性を確認する．

目 次

第 1章
序論

本研究は，秩序創発特性を有する問題解決システムの構築に関して考察を行なったものである．

研究背景

近年，我々の対象とする問題や取り巻く環境の複雑・大規模化に伴い，従来の中央集権型システムの限界が指摘されており，自律分散型システムあるいは創発的システムの設計に注目が集まっている．例えば，中央集権型システムの代表例ともいえる空港の航空管制システムも近い将来の航空交通網の大規模化に対して破綻を来たすであろうことが指摘されており，フリーフライトと呼ばれる自律分散型離着陸システムに関する研究が本格的に始められている．このようなシステムでは，従来の**トップダウン的**なシステム設計ではなく，秩序形成や創発性などの秩序創発特性を有する**ボトムアップ的**なアプローチをより積極的に活用したシステムの構築が求められている．

また，近年の社会環境の急激な変化および価値観の多様化に伴い，システム理論に課される役割も，組織化されたシステムを解析することではなく，環境の変化にも適応できる柔軟なシステムを創造することへ変わりつつある．すなわち，従来のシステム理論では，主として**定常**な環境における**線形**なシステムを対象としていたのに対して，環境変動もありかつ**非線形**なシステムである**複雑適応系**を扱う必要性が求められている．従来のシステム理論では，トップダウン的に対象を要素に分解し，その要素の特性を解明するが，対象の構成要素間の相互作用が弱くかつ構成要素の特性が線形である場合に有効であったが，非線形な相互作用が強い生物システムなどの複雑適応系には対応できない．

問題解決

問題解決は，システム工学や人工知能，認知科学などの分野で盛んに用いられてきた言葉である．**人工知能**の分野においては，問題が明確に定義され，目標や解決のために取りうる

手段が与えられている場合に，探索によって適切な手段を選択しながら目標に達する解を見出すこととして捉えられる [1]．**認知科学**でも，現在の状態に問題が生じていてそれが目標に達すれば解決できると期待されるときに，目標状態に指向して生ずる心的活動であると捉えられるとともに，人間の問題解決プロセスは，1) 問題を「理解する」プロセス, 2) 問題を「解く」プロセス, 3) 解を「吟味する」プロセスの三つに分けて考えられることが多い [2]．一方，**制御工学**においては，目標（制御仕様）から要請される望ましい応答を示すように制御システムを調整するメカニズム（制御システム）をいかに設計するかということであるが，近年の知的制御システムでは，制御とは対象となるシステムがある目的を達成するように適切な操作を加えるための問題解決を行なう一連の情報処理過程として捉えられる [3]．また，**システム工学**においては，問題をシステムすなわち複数の要素が相互に連関するものと認識し，このような認識対象を分析し，解決案に基づくシステムを構築し，運用・評価することとして捉えられる [4]．すなわち，互いに競合する多様な課題をどのように協調させるか，さまざまな事柄が関与する現実世界でのシステムの設計や制御・運用・診断はどうあるべきか，といった意味が含まれている．ここで，システム工学の方法論では，一般に問題解決の手順は以下のようにトップダウンに定式化される [5]．

ステップ 1　問題の定義（要求の明確化）
ステップ 2　目標の選択（評価基準の明確化）
ステップ 3　代替案の創成
ステップ 4　代替案の評価
ステップ 5　代替案の選択
ステップ 6　プロトタイプシステムの製作と実施
ステップ 7　本システムの実施と改善

ここで，問題というのは，望ましい状況と現状との差として定義される．システムは，この問題を解決する，すなわち現状を望ましい状況に変換するものとして導入される．このような目的のためには，システムはある特定の機能を果たさなければならず，機能を操作的に（達成度が定量的に測定できるように）定義したのが，ステップ 2 である．また与えられた機能を実現するシステムを考え出すのがステップ 3 である．一般に必要な機能を実現するシステムは唯一ではなく，技術的あるいは経済的，社会的な要因まで考慮して複数のシステムの比較を行なうのがステップ 4 である．ステップ 5 では，その評価のなかで最適なシステムを選択する．ステップ 6，ステップ 7 については文字通りの意味である．

問題の分類とモデリング

ここで，対象とする問題をその複雑さによって大別すると，以下の三つのレベルに分類することができる [6]．

1. *well-defined problem*: 従来の工学的な手法でモデル化でき，理論的・数値的に取り扱える問題．例えば，微分方程式系によってモデル化・表現できるような動的な対象システムなどであり，**良構造問題**とも呼ばれる．
2. *poor-defined problem*: 対象システムの振る舞いが定性的にはつかめるが，明確に数式モデルで表現するのが困難な問題．これらに対しては，問題設定からスタートしなければならず，また必ずしも有効な方法論が存在するとは限らない．
3. *ill-defined problem*: 対象システムの特徴が容易に把握できずいかなるモデル化も困難な問題．問題の構造化もなされていないので**悪構造問題**とも呼ばれる．

上記の第一の分類では，問題設定後の問題解決アルゴリズムが研究の中心であり，人工知能における古典的な問題や認知心理学で取り上げられる算数や理科の問題解決のなどがそれに入る．しかし，大規模複雑化するシステムの設計・制御・運用を考える際には，モデル化そのものが困難でできていない第二や第三の分類の問題について検討する必要があり，本論文ではそれらを主な対象とする．

対象問題の変化

このような構造化がなされていない悪構造問題などの問題に対しては，上で述べた従来のシステム工学・システム理論の**トップダウン的**な問題解決の方法では，対処することがますます困難になってきている．すなわち，問題が明確に与えられており，その解決の方式も明確であれば，問題を要素に分割し，それぞれの解決に適した能力・機能を有するシステム要素に割り当てるトップダウン的な解決がとれるが，問題が明確でなく構造化もなされていない悪構造問題に対しては，柔軟な**ボトムアップ的**な処理が問題解決過程に必要となる．したがって，秩序形成や創発性などの秩序創発特性を有するボトムアップ的なアプローチをより積極的に活用した問題解決の方法が強く求められている．

問題解決とシステム論

ここで，秩序形成や創発性などの秩序創発のシステム構成概念に基づく問題解決アプローチでは，まず対象問題そのものをシステム論的な視点から捉えることが重要である．対象と

する問題によってシステムとしての具体的な捉え方はそれぞれ異なるとは思われるが，例えば本研究の前半で述べるように，制約指向の観点から対象問題を捉えることはその一例といえる．これは，システム工学や制御工学などにおいて人間がトップダウン的に対象問題（対象システム）を捉える点で共通しており，**問題レベル**のシステム論といえる．

さらに，このような形で把握された対象問題の解決を図るプロセスとしての問題解決システムは，ある意味で視点を一段上げたメタレベルで問題解決を捉えていることになり，**メタレベル**でのシステム論ともいえる．すなわち，問題解決はシステムとして捉えた対象問題と問題解決を行なうシステムとの絶え間ない相互作用プロセスを通してなされるものと考える．したがって，問題解決そのものが静的なものではなく，動的な求解プロセスを重視したプロセス指向型の問題解決ということができる．例えば，本研究の前半で述べる自律的秩序形成プロセスや制約伝播プロセスによる制約指向問題解決および，後半で述べる強化学習の枠組みにおけるシステム－環境間の相互作用を介した行動学習プロセスは，その例として考えられる．また，そこでの重要な概念である秩序形成や創発性などの秩序創発特性は，問題レベルではなくメタレベルでのシステム構成概念であるといえる．

以下では，このような秩序形成や創発性などの秩序創発のシステム構成概念に関して提案されてきた理論や概念を概観し，秩序形成および創発性に関するシステム論の変遷の軌跡をたどる．

秩序化のシステム理論

ウィーナーは，**サイバネティクス** [7] の提唱において，開放系としての秩序を形成し，情報を生成する反エントロピー的な存在としての生物を強調した．つまり，生物システムは情報を収集し，それを加工しフィードバックすることにより，学習・適応を通してエントロピーを減じる（秩序を形成する）存在であると主張している．

ベルタランフィもまた，**一般システム理論** [8] の中で，生命・生物系を扱う理論の重要性を指摘している．生命・生物系の特徴は，ボトムアップ的なものであり，しかも組織化・秩序化される．要素還元的な立場では，システムの全体性や合目的性は見えてこないと述べている．

プリゴジンは，非平衡開放系にある非線形過程において，新しい特異点に到達したときに動的秩序が形成されるという**散逸構造** [9] の理論（非平衡系の熱力学）を熱力学第二法則の開放系への適用に基づいて提唱している．

ハーケンは，協同現象の数理とも訳される**シナジェティクス** [10] を提唱している．すなわち，システムの要素群が協調することにより，マクロな秩序を形成する過程を論じている．このような秩序形成は自己組織化とも呼ばれ，要素間の非線形な相互作用が本質的である．システムは，パラメータや状態の変化により，相転移や分岐現象として，ある時点を境に突然その状況を変容する．

創発性のシステム理論

ポランニーは，生命・生物系を対象としたとき，ミクロである下位レベルの作用により生じるマクロの上位レベルの動作の関係において，上位レベルの動作は下位レベルの法則によっては説明できず，そして上位レベルは下位レベルに対して境界条件を設定するものであるとし，これを**周縁制御の原理** [11] と呼んだ．例えば，機械の構成要素については物理・化学法則により説明できるが，上位レベルの特性である機械として作動原理は人間の与える合目的性により定まるものであり，下位レベルの法則からは出てこないということである．そして，生命・生物系においては，この上位レベルは創発により生む出されると述べている．

人工生命 [12, 13] の提唱者であるラングトンも，その中心的なキーワードとして**創発**を挙げている．人工生命では，次章でも述べるように，生命特有の現象をコンピュータ・シミュレーションなどで実現しようとしているが，このような生命の形態も存在し得ることはウィーナーのサイバネティクスで指摘されていたとも考えられるので，サイバネティクスの再来ともいわれる．

また，我が国では近年，文部省科学研究費重点領域研究「**創発システム**」[14] において，上記のような事柄を背景に創発性を指導概念として，生命・生物系の重要な機能である進化および適応を操作概念に置き，生物の創発的な進化と学習の過程を工学的に理解するとともに，生物的人工物を設計するための原理を明らかにすることを目指している．

本研究の目的

以上のような背景のもと，本研究においては，“秩序創発”を自律的秩序形成および創発的秩序形成の二つに大きく分けて考え，それぞれの原理に基づくシステムによる「問題解決」に関する考察を行なう．すなわち，前者に基づくシステムを「システム内部における秩序形成による問題解決」のためのシステムとして，後者に基づくシステムを「システム－環境間における秩序形成による問題解決」のためのシステムとして提案し，設計問題や制御問題などのさまざまな問題に対する適用を通して提案システムの有効性を明らかにする．

本論文の構成

以下，本論文の各章の構成について述べる．

第 2 章においては，本研究の背景として，秩序形成と創発性の概念について述べるとともに，自律的秩序形成に基づく問題解決および創発的秩序形成に基づく問題解決に大きく分けて考える．前者は「システム内部における秩序形成による問題解決」として，後者は「システム－環境間における秩序形成による問題解決」として位置づけて議論する．さらに，関連する研究分野についてもいくつか取り上げて紹介する．

第 3 章においては，自律的秩序形成による問題解決システムとして，二階層から構成される制約指向型自律分散システムを提案し，連続変量・ファジィ制約を含む制約充足問題への適用について述べる．

第 4 章においては，第 3 章でのアプローチとは異なる制約指向型の解法として制約伝播に着目し，制約指向問題解決プロセスに潜む複雑性について明らかにする．

これら 2 つの章は，「システム内部における秩序形成による問題解決」に関して考察したものといえる．

第 5 章においては，適応の機構を組み込んだ創発型システムの実現例として，ファジィ推論を導入することにより連続値入出力を扱う強化学習について提案し，制御問題への適用を通してその有効性を確認する．

第 6 章においては，学習の効率化・高速化を図るために経験強化を考慮した強化学習について提案し，自律移動ロボットを含むいくつかの制御問題への適用を通してその有効性を確認する．

これら 2 つの章では，「システム－環境間における秩序形成による問題解決」に関して考察しているといえる．

最後に，**第 7 章**では，本研究全体の結論として，各章において得られた成果をまとめ，それらについて検討を行なう．また，その際に現れた問題や今後に残された課題に関して考察を行なう．

第 2 章
研究の背景

2.1 緒言

情報化・ネットワーク化の急速な進展および価値観の多様化に伴い，我々が対象とする問題はますます大規模かつ複雑化・多様化しており，複雑な環境の変化や多様な要求に適応できる柔軟性・拡張性などが求められている．この場合，従来のシステムのように，入出力関係を明確に記述し，それに基づいて設計者が最適な解を求め，システムに組み込むといった手法では対処が困難になってきている．それに対して，脳・神経系，免疫系，生態系などの生物システムでは，構成要素の協調によって大域的な秩序として機能／構造が生み出され，それが環境の変化とともに更新される特性である“秩序形成”と“創発性”といった**秩序創発**のメカニズムが備わっている [14]．

本章では，本研究の基礎概念である秩序創発を**自律的秩序形成**と**創発的秩序形成**の二つに大きく分けて考え，それぞれの概念に基づく問題解決について概説する．まず，秩序形成と創発性の概念について，システムの階層性や環境との関わりに関連づけて述べるとともに，自律的秩序形成に基づく問題解決および創発的秩序形成に基づく問題解決に大きく分けて考え，それぞれを「システム内部における秩序形成による問題解決」と「システム－環境間における秩序形成による問題解決」の二つに対応させて議論する．つぎに，自律的秩序形成に基づく問題解決について，自己組織化の原理を説明し，それらに基づいた制約指向型自律分散システムを考える．さらにこの背後にある制約指向問題解決プロセスに潜む複雑性についても言及する．つぎに，創発的秩序形成に基づく問題解決について，環境との相互作用を重視した立場に基づき，進化および学習のメカニズムを導入した創発型問題解決システムについて考える．

さらに，以上のような自律的秩序形成および創発的秩序形成に基づいた問題解決に関連した研究分野として，人工生命や複雑系，ロボット学習，マルチエージェントシステムなどを取り上げ，簡単に紹介する．

2.2 自律的秩序形成と創発的秩序形成

2.2.1 秩序形成（自律的秩序形成）

システムの複雑・大規模化に伴い，従来の一極集中型のシステムでは柔軟性や拡張性に問題があることが指摘され，サブシステムと上位システムの疎に結合された階層構造を活用し，増大する複雑さに対処するシステム概念として**自律分散システム** (Autonomous Decentralized Systems)[15] が提案されている．ここで，自律分散システムとは「システム全体を統合する機構を持たず，分散して存在するサブシステム群から構成され，各サブシステムが自律的に行動しながら互いに協調し，システム全体としての包括的目標を達成する（秩序を形成する）システム」である [16].

このような自律分散システムの構成において不可欠なものは，各サブシステムがシステムの包括的目標に向けて協調し，全体的な秩序を形成する**自律的秩序形成**の機能である．すなわち，自律分散システムは各サブシステムの自律的問題解決活動とそれらのシステム全体への協調的統合可能性によって高い柔軟性を追求する試みであり，協調の結果としてシステム全体に形成される秩序が問題解決につながるものとなる．これによって，効率性・信頼性・拡張性・柔軟性の高い問題解決システムの実現が期待されるが，実際には概念が先行し，具体的なシステムの構築は少数にとどまっているのが現状である．

ところで，さまざまなシステムが階層的な構成の下で全体の秩序を維持していることは，生物学や社会科学の分野でしばしば言及されてきた．このようなシステム構造のもつ階層性は，自律分散システムの構成において最も重要である．例えば，生物システムである動物を例に階層性を見ると，図 2.1のようになる．ここで，サイモンによれば，多くのシステムが階層性をもつ理由は，増大する複雑性への対処，とくに進化と環境変化への適応の容易さにあり，階層構造システムの方が非階層システムに対して優位であることを指摘している [17]. また，ケストラーは，ホロン (holon) と呼ばれる単位からなる階層構造によるシステム構成について提唱している [18]．ここで，ホロンとは，上位構造に対する部分性と下位構造に対する全体性の二つの側面（二面性）を有する問題解決単位である．

このような階層構造を支える統御原理として注目すべきものに，ポランニーにより提唱された**二重制御** (dual control)[11] の概念がある．これは，各階層それぞれが，下位の階層を構成する諸要素を支配する法則と，自分自身の階層の諸要素のものに対して成り立つ法則に同時に（二重）に制御されているということである．例えば，動物の個体はその構成要素である器官を支配する法則によって制御される．すなわち動物はその構成要素である各種器官のもつ機能の働きからのがれることはない．それと同時に，個体はその個体自身を支配する

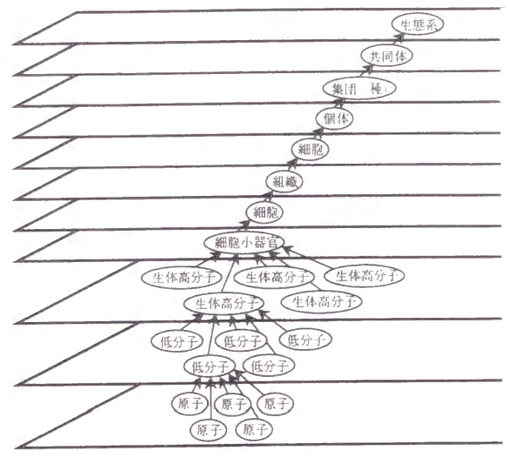


図 2.1: 生物システムの階層性

法則にも制御される．つまり，個体はその個体自身の行うことのできる行動や精神活動の枠を超えることはできない．このように個体は二重制御を受けていると考えられる．

このような各階層が受ける二重制御は，**周縁制御の原理** (principle of marginal control) に従っている．つまり，下位の階層は自身の諸構成要素を支配する法則に従う一方，上位の階層の組織原理は下位の階層の構成要素間の境界条件を決定する．例えば，動物の個体の階層を上位，器官の階層を下位とする二階層を考えると，筋肉その他の諸器官の機能は個体の運動が成り立ち得る可能性を開いており，そのような確定されないまま残されている境界条件を上位の個体の運動の組織原理が決定する．

2.2.2 創発性（創発的秩序形成）

創発 (emergence) という言葉は，生物進化に関して提唱されたといわれている．哲学辞典では，創発とは「進化論で用いられる概念で，先行与件から予言したり，説明したりすることが不可能な進化・発展をいう」とある．つまり，生物進化の長い歴史において，遺伝情報（遺伝子型）から環境との相互作用で発生・形成される生命（表現型）などが想定される．また，近年の**創発システム** (Emergent Systems)[14] に関する研究では，創発の定義は以下のように与えられている（図 2.2参照）．

「システム構成要素間および環境との局所的な相互作用を通じて，大域的な秩序がボトムアップ的に発現し，こうしてできる大域的な秩序が境界条件として要素間の局所的相互作用をトップダウン的に支配する双方向の動的過程を通して，新しい機能・形質・行動などの獲

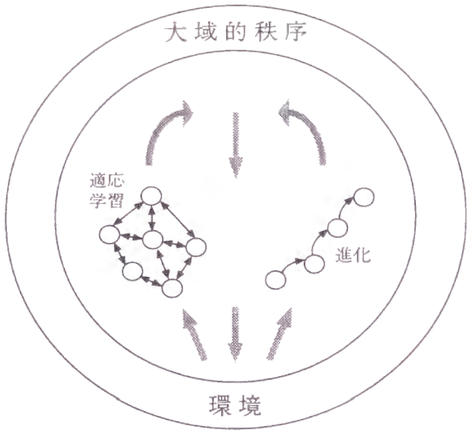


図 2.2: 創発の概念（文献 [14] より引用）

得をもたらすこと」

このような創発の定義は，人工生命における創発の定義とほぼ同じものであり，ポランニーのいう周縁制御の原理や，ハーケンの提唱するシナジェティクスとも共通点をもつ捉え方である．本研究では，このような創発性に基づく秩序形成・機能形成を**創発的秩序形成**と呼ぶ．

なお，創発の概念そのものについては，このほかにも実にさまざまな捉え方が歴史的になされてきた [21]．例えば，最初に「創発」という用語を使ったといわれる 19 世紀のルイスは，生命現象と人間の精神活動を説明するのにこの概念を用いた．今世紀初めには，モルガンが生物の進化における創発現象（新しい形質や特徴が出現すること）を強調するとともに，全体論と関連づけて哲学的に論じている．また，生態学のウィルソンは，生物個体と社会との双方向の創発関係を基本に創発的特性について議論している．このウィルソンの創発論は，人工生命のラングトンの創発概念につながるものといわれている．

上記のような創発性を有するシステムである創発システムにおいては，システムの機能や構造を外部の設計者が与えるのではなく，創発的にその機能や構造を発現・形成させることが重要である．また，このような創発システムを実現するためには，創発性を実際に示している生物システムに学ぶべきであり，とくに環境との相互作用を通した長期間にわたる進化の機能と短時間での適応・学習の機能に注目することが重要である．

本研究では，このような秩序創発特性をシステムと環境の関わりの観点から以下の二つに分類し，以下の議論を展開してゆく．

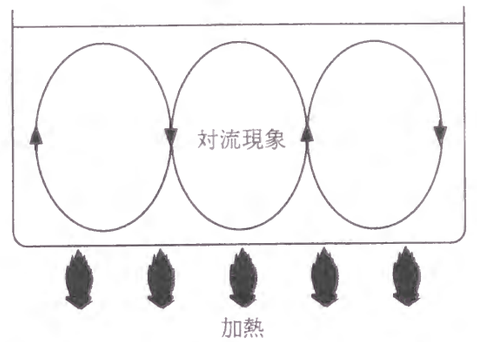


図 2.3: 対流における動的秩序の形成

(a) システム内部における秩序形成

秩序形成・自己組織化の原理に基づく自律分散システム，自律的秩序形成に基づく問題解決に対応する．

(b) システム−環境間における秩序形成

環境適応のための進化・学習能力を有する創発システム，創発的秩序形成に基づく問題解決に対応する．

なお，次の **2.3** 節では (a) に対応する自律的秩序形成と問題解決について述べ，つづく **2.4** 節では (b) に対応する創発的秩序形成と問題解決について議論する．

2.3 自律的秩序形成と問題解決

2.3.1 自己組織化の原理

自己組織化という概念は古くから現在に至るまで様々な意味で用いられている．古くはウィーナーのサイバネティクスに始まり，アイゲンやプリゴジン，ハーケン，清水博などが少しずつ異なる意味でこの言葉を使っている．ここでは，物理化学的な自己組織化現象の追求として，プリゴジンの**散逸構造** (dissipative structure) の理論とハーケンの**シナジェティクス** (Synergetics) を取り上げる．

散逸構造（非平衡系の熱力学）

散逸構造の理論は，非平衡にある非線形過程において，新しい特異点に到達したときに動的秩序が形成されるというもので，水を熱したときに見られる対流が例である（図 2.3参照）．

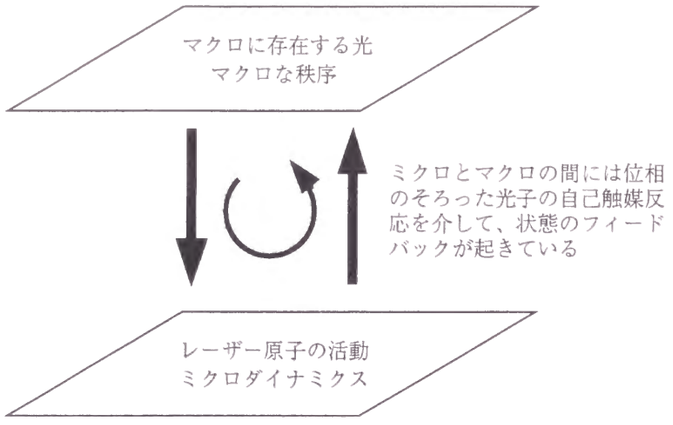


図 2.4: スレイビング原理成立の模式図

容器の上部と下部の温度差が小さい間は，下部から上部への熱エネルギーの輸送メカニズムは分子のランダム運動である．つまりランダム運動の際に起こる分子の衝突により，分子から分子へ運動エネルギーが伝えられる．しかし上部と下部の温度差が大きくなると，このような形の熱エネルギーの輸送が下からの熱エネルギーの供給に間に合わなくなり，容器の上部には冷たい（密度の高い）液体が，また下部には暖かい（密度の低い）液体がたまった力学的に不安定な構造となる．その後，分子のゆらぎが引き金になって粘性の支えを破ることで不安定構造がひっくり返り，対流というマクロな秩序が形成される．この対流のように，エネルギーを使うことによって生み出される動的な秩序構造を散逸構造という．

シナジェティクス（協同現象の数理）

ハーケンのシナジェティクスは，物理現象のみならず経済や社会に至る幅広い範囲の自己組織化現象を統一的に扱うもので，部分系の協同作用がどのようにして巨視的スケールの空間的，時間的あるいは機能的な構造をもたらすかを研究するものである．そこでの考え方の核心は，スレイビング原理（隷属原理）と呼ばれる原理により秩序が自己組織化されるメカニズムを説明づけることである．

ハーケンは，自己組織化現象の簡単な例として，レーザーを取り上げている．レーザー管の中でポンピングと呼ばれる操作によって励起状態に置かれた原子は，もとの基底状態に戻る際に誘導放出によって光を放出する．この場合，外からの光と同じ位相の光が放出される．ポンピングによって与えられるエネルギーが大きくなると，励起状態の原子の密度の高い不安定構造となり，誘導放出が連鎖的に起こり，原子の状態変化が協同的に進行する．その結

果，レーザー管の両端に取り付けられたミラーの働きで，レーザー管の軸方向に位相の揃ったレーザー光（マクロな秩序）が発生する．このとき，誘導放出の働きで，原子のミクロな運動が，マクロな秩序であるレーザー光に同調することになる．この誘導放出は，マクロな秩序に自己のミクロダイナミクスを同調させる性質による，すなわち原子の運動のもつ**自己触媒性**に基づくものと解釈される．

位相の揃った光の波というマクロな秩序は，多数の原子の内部運動というミクロダイナミクスから構成されている．またその一方でミクロダイナミクスは，マクロな秩序に隷従している．つまり，マクロな秩序とミクロダイナミクスの間にはフィードバックループが存在する（図 2.4参照）．ハーケンは，秩序形成に関して，このような隷従化現象が起きているときに**スレイビング原理** (slaving principle) が成立するといい，レーザーばかりでなく幅広い領域での自己組織化現象において成立するものと主張している．

2.3.2 制約指向型自律分散システム

本節では，前節の自己組織化の原理と制約指向型の問題解決に基づく自律分散システムを考える．ここで，**制約指向問題解決**とは，制約という観点を中心に問題を捉えることにより，問題を制約により記述し解決を図るアプローチである．

この制約指向の立場に基づき，制約（制約条件，制約領域）により記述した複雑な問題を**ファジィネス**の導入によって，緩やかに結合した（比較的独立な）部分問題にいったん**分解・還元**し，それぞれの部分解を求めた後に，それらを再び**合成・統合**することによって問題解決を行うことを考える．

本研究では，自律分散システムの構成原理に関する考察に基づき，記号処理を行う上位層と集合力学的計算を行う下位層の二層から構成される階層型自律分散システムを考える．すなわち，**記号処理に基づく上位層**が大局的・構造的な制約の処理を担当するのに対し，集合力学的なダイナミクスに基づく**自律分散型処理を行う下位層**が，局所的・詳細な制約を扱う．両層が互いに影響を及ぼしあいながら，全体としての秩序を自己組織化することにより，与えられた問題の解決が計られる．その際，シナジェティクスの基本原理であるスレイビング原理を参考に，上位層・下位層間に相互作用を設定する．これら二つの階層内の計算プロセスは，すべて同時並行的に実行される．

すなわち，本来あいまいさの内包されない制約充足問題に対しても，人為的にファジィネスを導入し，制約を分解・還元することによって，解の構造的選択－上位層での処理－と，解の連続的な許容範囲（制約区間）の選択－下位層での処理－に還元することが可能である

ことを示し，これらの選択を同時並行的かつ重畳した形で実行するシステムとして，全体の論理的整合性を図る記号処理と局所的な相互作用によって解全体のバランスをとる自律分散型処理という二つの計算原理を含む階層型の自律分散システムの枠組みを提案する．

このような階層型問題解決システムは，秩序形成原理・自己組織化原理を導入した自律分散システムであり，システム内部（上位層・下位層の間）における創発性を実現するシステムと捉えることができる．

次の**第 3 章**において，このような自律分散型処理と記号処理という二つの計算原理を含む階層型自律分散システムの枠組みを提案する．

2.3.3 問題解決と複雑性

前節のようにネットワーク内の全てのリンク構造を同時並行的に活用し，自己組織的・自律分散的に問題解決を行うのではなく，一度に一つのリンク構造だけを選びそれらを逐次的にたどる制約伝播による問題解決が考えられる．このとき，伝播されるのは制約区間であり，制約伝播の際に連続量を記号化する手段としてファジィネスが導入された系（ファジィ記号力学系と呼ぶ）の制約伝播構造にはカオス的現象を生み出すような複雑な構造が内包されていることを明らかにする．また，記号化のためのファジィネスを導入しない系（制約伝播力学系と呼ぶ）では，安定平衡点に収束する極めて安定的な振る舞いが得られることを示す．

このような制約指向問題解決に内在する複雑性および安定性について，**第 4 章**において明らかにする．

このような制約伝播プロセスによるアプローチでは，絶えず様々なパターンを次々に作り出しており，極めて多様な解（全体的整合性は満たされていないかもしれないが）を常に探索していると考えられる．そこで導出される解は，解自体が動的な意味をもっており，従来の収束解を包含した形での新しいプロセス指向の解概念と捉えることもできる．この意味で複雑性を内包する制約伝播による問題解決は，静的な解のみを追求するのではなく，動的な求解プロセスを重視したプロセス指向型の問題解決ということができる．

2.4 創発的秩序形成と問題解決

2.4.1 環境への適応性

自律分散システムと創発システムはともに，サブシステム群から構成され，それらが自己組織化して全体として高次の機能を発現・形成するという点では共通である．しかし，自律分散システムの自己組織化が，ある与えられた環境において動作しているうちに秩序形成が

行われ，ある安定状態に至ることであるのに対して，創発は，環境が変化した際に，すでに形成された秩序が壊れ，過渡的な状態を経て，別の秩序が形成されたときにおける環境変化に対する適応性であると考えることができる．

創発的なシステムを人工的に構成するためには，システムと環境との関わりに着目することが非常に重要である．すなわち，環境という場がこれまで以上に重要な役割を果たすと考え，問題解決をシステムと環境との絶え間のない相互作用による動的プロセスとして捉えることが不可欠である．そのうえで，システムから環境への働きかけと環境からの反応を介して，より環境に適合したシステムを自己組織的に実現する方法として，適応的学習および進化プロセスの導入が考えられる．

2.4.2 進化・学習能力

創発的なシステムの実現に向けて，生物システムの自己改善メカニズムとしての進化および学習の能力に注目し，システムの環境への自律的な適応の実現を図る．環境への適応という観点から見ると，進化は長期的な適応であるのに対して，学習は短期的な適応に対応する．このような進化・学習能力を実現する方法論として，**進化型計算**と**適応的学習**があり，それらは創発システムの重要な要素技術として位置づけられる．ここでは，両者の基本的な枠組みについて以下に概観する．

進化型計算

進化型計算とは，生物の進化過程を模倣した計算論的枠組みの総称概念であり，**遺伝的アルゴリズム** (Genetic Algorithm)，**遺伝的プログラミング** (Genetic Programming)，**進化的戦略** (Evolutionary Strategy)，**進化的プログラミング** (Evolutionary Programming) などがある．

進化型計算の全ての枠組みに共通するメインループは，選択 (selection) と交叉 (crossover)・突然変異 (mutation) であるが，それぞれが固有の表現と方法を採用している．表現形式としては，遺伝的アルゴリズムでは文字列，遺伝的プログラミングでは木構造，進化的戦略と進化的プログラミングでは実数値が用いられる．遺伝的アルゴリズムは交叉を主たる探索オペレータとするのに対し，進化的プログラミングでは突然変異のみを探索オペレータとしており，進化的戦略では交叉と突然変異の両方を採用している．

ここでは，遺伝的アルゴリズムの枠組みについて簡単に述べる．生物の進化過程においては，各個体が有する遺伝子が非常に重要な役割を果たす．まず，親の遺伝子が複製されて子

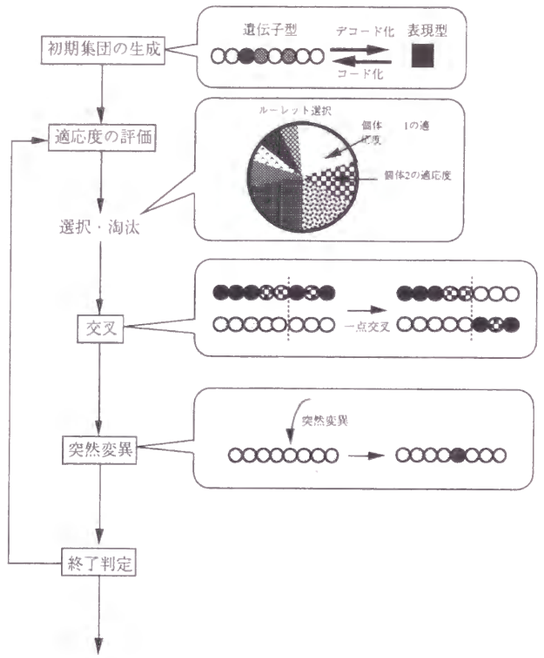


図 2.5: 遺伝的アルゴリズムの枠組み

の遺伝子が生成される際に，交叉や突然変異がなされ，親とは異なった形質を発現する．こうしてできた表現型が，環境への適応度という評価により淘汰・選択され，適応度の高い優秀な個体群が次世代に引き継がれる（図 2.5参照）．

このような遺伝的アルゴリズムによる探索の特徴は，1) 基本的に多点探索であり局所解に陥ることが少ない（大域的な探索能力が高い），2) 評価値の勾配などを用いないので不連続な評価関数の探索に適している，ことなどが挙げられる．

適応的学習

生物進化の根幹をなす機能の一つである適応・学習機能にそのアナロジーを求め，人工システムに適応・学習機能を付与することを目指した研究分野であり，環境との相互作用を通して得られる報酬 (reward) のみを手掛かりに環境に適した行動を自律的に獲得する**強化学習** (Reinforcement Learning) はその代表的な枠組みである．

強化学習においては，学習主体であるエージェントは環境の状態をセンサにより認識した後，行動を選択し実行する．その際，一連の行動に対して環境からの報酬や罰が強化信号として与えられ，学習エージェントはより多くの報酬の獲得する行動を強化してゆく（図 2.6

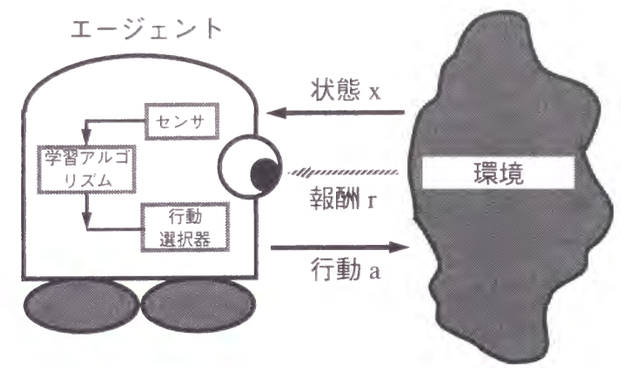


図 2.6: 強化学習の枠組み

参照）．

強化学習の主な特徴としては，1) 基本的には教師なし学習であり，試行錯誤的な探索を通して学習がなされる，2) 環境からの評価が行動系列に対して遅れて与えられる遅れ報酬系である，点などが挙げられる．

機械学習のサブクラスである強化学習は，2 つのアプローチに分けて議論されることが多い．1 つはそれまでの経験を重視する**経験強化型アプローチ**であり，分類子システム (Classifier System) に代表され，最適性は保証されないがその収束性の速さで知られる．他方は環境をより広く探索する**環境同定型アプローチ**であり，Q-Learning が代表的であり，環境の状態遷移にマルコフ性が仮定すると最適値への収束が保証されるが学習コストは大きい．

2.4.3 創発型問題解決システム

すでに述べたように，システムから環境への働きかけと環境からの反応を介して環境に適応することにより問題解決を行なう創発型システムを実現する方法として，適応的学習および進化プロセスの導入が考えられる．本研究ではとくに，システムと環境との絶え間のない相互作用に基づいて，システムを環境に適合したものに自己組織的に改良してゆくことを，強化学習などの適応的学習の理論をベースに実現することを検討する．

具体的には，環境同定型の強化学習アルゴリズムである Q-Learning に注目し，ファジィ推論を導入することにより連続値入出力を扱える強化学習法の提案を**第 5 章**において行う．また，遺伝的アルゴリズムを用いた機械学習法である分類子システム（Classifier System）で提案された Profit Sharing 法の考え方を Q-Learning に対して導入した手法を**第 6 章**において提案する．

このような手法を実現したシステムは，システムと環境との絶え間のない相互作用による動的過程に基づく適応プロセスであり，システムと環境の間における創発性を実現するシステムと捉えることができる．

また，自然の生物システムは，本質的に豊かな多様性・冗長性・複雑性を有しており，それらが環境変化に対応するための源となっている．上述のような創発的システムにおいては，従来の工学的システムでは否定的に捉えられていたこのような多様性・冗長性・複雑性を積極的に有し活用することが極めて重要であると考えられる．

2.5 関連分野

2.5.1 人工生命

人工生命とは，自然界の生命に特有な現象・挙動をコンピュータ・シミュレーションや人工的なシステムを作ることより実現し，生命の本質に迫ろうとする研究である [12]．具体的な手法としては，セルオートマトンや遺伝的アルゴリズム，ニューラルネットワークなどが用いられることが多いが，それらに共通する思想的背景はやはり本章で述べた**創発**の概念である．ただし，人工生命の解釈や研究内容は多様であり，生物の細胞レベルの発生過程から，樹木の成長などの個体の発生過程をシミュレーションするものもあれば，生物の進化や個体の行動・群れの行動などをシミュレーションするものもある．

例えば，アリの集団的集餌行動をコンピュータ上に模擬・実現した例 [13] がある．この例では以下のような単純なモデルが仮定される．アリは，うろうろと餌を求めて歩いており，餌が見つかるとフェロモンを分泌するとともに餌を持って巣へ向かう．すると餌から巣へ向かってフェロモンの跡ができ，他のうろつき歩いているアリは，フェロモンの濃度が濃くなっている方向（餌のある方向）へ引き付けられ，餌を発見することができ，最初のアリと同じくフェロモンを分泌して巣へ向かう．こうしてフェロモンの跡はますます濃くなり，さらに多くのアリが引き付けられる．しかし餌はいつかはなくなり，フェロモンもすぐに蒸発して消えるので，アリの集団は再びうろうろと餌を求めて歩き回る．このアリの行動は下位の局所的行動であり，それが自発的にフェロモンの濃度という大域的秩序を形成する．その秩序は各アリの行動に影響を及ぼし，餌を運ぶという集団運動が出現する．このように双方向の規定関係である創発が実現されている．

2.5.2 複雑系

複雑系とは，従来の還元主義的な自然科学では扱えない様々な非線形現象を対象とする学問領域である [27]．自然界における非線形現象は，線香からたなびく煙の動きなどのように身近に存在する現象から世界的な規模の気象の変化などにいたるまで実に様々なものがある．複雑性を扱う数学的な基盤としては，**カオス理論** [28, 29] の分野がこれまで精力的に研究されている．

例えば，カオスを生み出す要素が多数結合された**結合カオス系** [30] というものがある．結合にはいろいろなタイプがあるが，ここではまったく同じカオスを示す要素があり，それぞれの平均値を通して全部が影響を受ける系（大域結合カオス系）を取り上げる．この系の基本的なパラメータとしては，個々の要素のカオスの強さを与えるものと，全体の平均との結合の強さを与えるもの２つがある．平均との結合が大きければ，要素は引き込んで振動するコヒーレントな状態になり，カオスが強ければ要素の振動は完全にばらばらになる．さらに，この２つの中間の状態として，全要素は引き込んで振動する幾つかの集団（クラスター）に分かれ，それぞれの集団では揃って振動するようになる．つまり，まったく同じ要素の集団であったのに違った振動をする集団に分化したのであり，カオスの結合が多様性を生むもとをなし得るといえる．

2.5.3 ロボット学習

ロボティクス，人工知能 (AI) の分野で環境との相互作用を重視したアプローチに，Brooks により提唱された**行動型 AI** がある．Brooks は，従来の AI での知識表現や推論を否定し，*Situatedness*（システムが現実の状況に信号的に直結していること）と *Embodiment*（システムが実世界に物理的行動を通してつながっていること）を重要視し，具体的には階層並列化された多数の行動モジュール群からなる**サブサンプション・アーキテクチャ**[31] を提案し，様々なロボットへの実装を通して有効性を示した．

このような行動型 AI に対して，学習能力を付与するための研究が近年盛んに行われている．手法としては，ニューラルネットワーク，進化型計算，強化学習などの手法が用いられることが多い．ここでは，ニューラルネットワークを用いて力学系に基づく行動学習の手法を自律移動ロボットに実現した例 [32] について取り上げる．この例では，ロボットが目的のタスクを達成する過程を作業座標系上のあるアトラクタに収束するダイナミクスとして捉え，この軌道を埋め込む適切な内部空間をセンサ情報をもとに構成する．内部状態からモーターコマンドへの写像は，ニューラルネットワークを用いた教師付き学習を通して獲得され，

センサ情報に基づいたナビゲーションが実現できる．

2.5.4 マルチエージェントシステム

マルチエージェントシステムは，分散人工知能の一分野であり，複数の行動主体（エージェント）が協力し全体の問題の解決を行なうシステムである．分散人工知能は，共通の目標の達成を目指す**協調問題解決**と各エージェントが独立の目標をもつ**交渉・均衡化**に分類されることがある [33] が，マルチエージェントシステムはどちらかというと共通の目標を有することが多いと思われる．

このようなマルチエージェントシステムに期待される点としては，1) 単体では不可能だった作業が多数のエージェントの協調動作により可能となる，2) あるエージェントが故障しても他が肩代わりできるのでシステムの信頼性が向上する，3) 多数のエージェントが協力して問題の解決に当たるため全体の作業効率が向上する，ことなどが挙げられる．

また，本研究で扱う自律分散システムとマルチエージェントシステムと自律分散システムの違いは，後者が単独でも知的処理が可能な高度な主体から構成される，いわゆる粒度の粗いシステムであるのに対して，前者は個体としては比較的単純な要素を多数集めて構成される粒度の細かいシステムであるといえる．

マルチエージェントシステムにおける自己組織化の例としては，与えられる問題の状況に応じてエージェント組織の合併や分割などの再編を自律的に行なう例がある．すなわち，エージェントは問題解決前にタスクを割当てるだけでなく，問題解決中にも必要に応じて組織を再編することが求められる．このような自己組織化は，複数のエージェントが問題解決タスクを共有し，サブタスクへの分割を通じて負荷を分散する協調の形式であるタスク共有の一般化となっている．ただし，ここでの自己組織化 (Organization Self-Design) は，自律分散システムや創発システムでの自己組織化 (Self-Organization) とは少し意味合いが異なる．

2.6 結言

本章ではまず，秩序形成と創発性の概念について，システムの階層性や環境との関わりに関連づけて述べるとともに，自律的秩序形成に基づく問題解決および創発的秩序形成に基づく問題解決に大きく分けて考え，それぞれを「システム内部における秩序形成による問題解決」と「システム－環境間における秩序形成による問題解決」として捉えることを述べた．

自律的秩序形成に基づく問題解決については，2.3 節において，自己組織化の原理を概説

した後，それらに基づいた制約指向型自律分散システムの枠組みについて述べた．このシステムの詳細については，第 3 章において提案する．さらにこの背後にある制約指向問題解決プロセスに潜む複雑性についても言及したが，その複雑性に関しては第 4 章で計算機シミュレーション等を通してより具体的に明らかにする．

創発的秩序形成に基づく問題解決については，2.4 節において，環境との相互作用を重視した立場に基づき，進化および学習のメカニズムを導入した創発型問題解決システムについて考えた．その具体的な実現例としては，第 5 章および第 6 章において，強化学習を導入したシステムを提案する．

第 3 章
自律的秩序形成による問題解決

3.1 緒言

近年，我々が取り扱う問題の規模や複雑性の増大に伴い，問題解決アルゴリズムの発見がますます困難になってきている．そのため，**制約に基づく問題解決**に多くの期待が寄せられている．そこでは，問題を制約すなわち対象要素間の関係構造によって規定し，すべての制約を満足する解を求める制約充足問題解決がなされる．

変量（変数）が離散値の**制約充足問題**に関しては，さまざまな解法が提案されてきたが [34, 35]，連続変量を含む制約充足問題の有効な解法は得られていない．また，あいまいさを含んだファジィな制約の扱いについても，十分議論されているとはいえない．また一般に，制約充足問題は NP 完全な問題であるため，計算量や記憶容量などの面で効率よく解を求める近似解法の導入が望まれる．

本章では，制約指向の立場からファジィネスを捉え，ファジィ情報処理の中に，全体の論理的整合性に注目した記号処理の部分と，局所的な相互作用を通して解全体のバランスをとる自律分散型処理が内包されていることを明らかにし，二つの計算原理を含む階層型の**自律分散型問題解決**の枠組 [42] の導入と，連続変量・ファジィ制約を含む制約充足問題への適用法を検討する．

そのために，我々が先に導入した**区間制約ファジィ集合概念** [39] を用いて，対象問題に内包される制約群を区間制約の集合体ならびに区間制約間の対応関係という，二種類の制約構造に分解する方法を導入したのち，自律分散システムの構成原理に関する考察に基づき，連続変量・ファジィ制約を含む制約充足問題の解決システムとして，二つの計算原理を含む階層型自律分散システムを提案する．記号処理に基づく上位層が大局的・構造的な制約の処理を担当するのに対し，集合力学的なダイナミクスに基づく自律分散型処理を行う下位層が，局所的・詳細な制約を扱う．両層が互いに影響を及ぼしあいながら，全体としての秩序を自己組織化することにより，与えられた問題の解決が計られる．その際，**シナジェティクス** [10] の基本原理であるスレイビング原理を参考に，上位層・下位層間に相互作用を設定す

る．これら二つの階層内の計算プロセスは，すべて同時並行的に実行される．

以下，**3.2節**において，制約充足問題の概要とその代表的解法について説明を行なった後，我々が先に導入した区間制約ファジィ集合概念を紹介する．さらに，区間制約ファジィ集合を用いることにより，連続変量・ファジィ制約を含む制約充足問題を整合ラベリング問題（CLP）に帰着できることを示す．**3.3節**においては，自律分散システムの構成原理に関する一般的な考察を行なった後，記号処理と自律分散型処理を融合した階層型の制約充足問題解決システムを提案する．さらに**3.4節**では，提案システムを並列処理言語 Occam[44] によりトランスピュータ上に実装し，設計問題への適用を通して，提案システムの有効性を確認する．

3.2 制約充足問題と制約指向型ファジィネス

3.2.1 制約充足問題とその定式化

制約充足問題とは，複数の構成要素からなる対象に対する解釈・合成・設計などを，付随する制約条件を充足する解を求める問題として捉えたもので，各構成要素間には局所的な制約条件が課せられ，さらに対象全体にも何らかの全体的整合性が求められる問題である．

変量が離散値の制約充足問題はとくに**整合ラベリング問題**（CLP : Consistent Labelling Problem）[34] と呼ばれることがある．これは，複数存在するユニット（変数）それぞれに，ユニット間に課せられた制約条件を満たすようなラベル付けを求める問題である．すなわち，ユニットの集合 $U = \{X_1, \ldots, X_n\}$，候補ラベルの集合 L，ユニットの多項組（ユニット拘束関係）の集合 $T = \{t_1, \ldots, t_m\}$，さらに各ユニット組に対してラベル付け可能なラベルの多項組の候補（ラベル拘束関係）$R_i = \{R_{i1}, \ldots, R_{im}\}$ の集合が与えられ，すべてのラベル拘束関係を満たすようなラベル付け（ラベルの割当て）を各ユニットに対して行う問題である．よく知られた例としては，クロスワードパズルやグラフ彩色問題，線画解釈問題などがある．

CLP の重要な変形として，候補ラベル組に重みをつけた**概整合ラベリング問題**がある．この重みは各候補ラベル組のそれぞれの確からしさを数値で表したものであり，その確からしさが強いほどその重み（誤差）が小さいものと考えることができる．したがって，これらの重みを集積した全体誤差を考えると，概整合ラベリング問題を解く方法としては，(1) 全体誤差がある閾値以下のものを求める，(2) 全体誤差が最小のものを求める，の二通りの方法がある．

CLP は，**制約ネットワーク**と呼ばれるグラフにより等価表現される．頂点はユニット（変数）を表し，頂点間の辺は頂点が表す変数間に制約が存在することを表す（図 3.1）．一般に CLP は，NP 完全な組み合わせ探索問題であり，処理の迅速性を図りつつ近似解を求めることが必要となる．

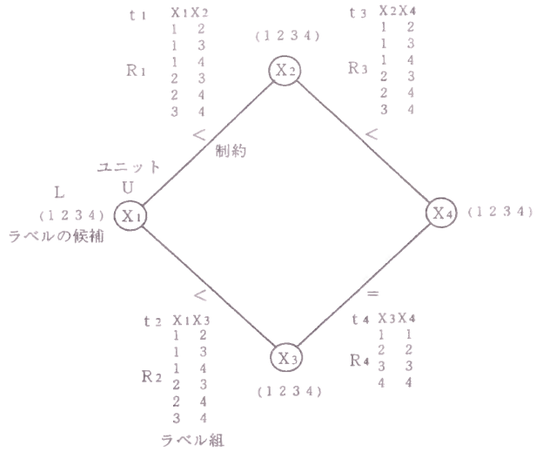


図 3.1: 制約ネットワークによる表現

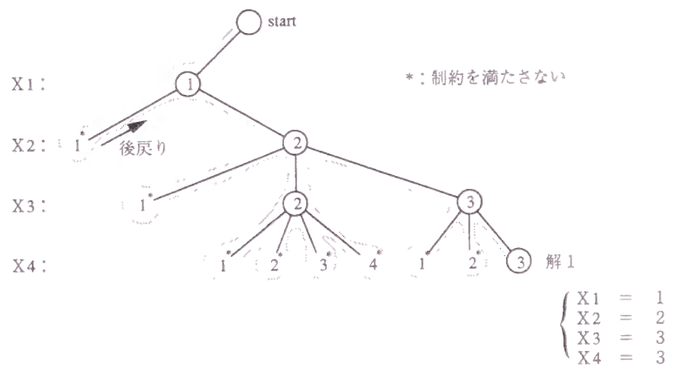


図 3.2: 木探索法による求解過程

3.2.2 制約充足問題の代表的解法

制約充足問題の解法としては，さまざまな手法が提案されているが，解構成型アルゴリズムと状態空間アルゴリズムに分類することができる [35]．本節では，この分類に基づいていくつかの代表的な解法について簡単に説明を述べる．

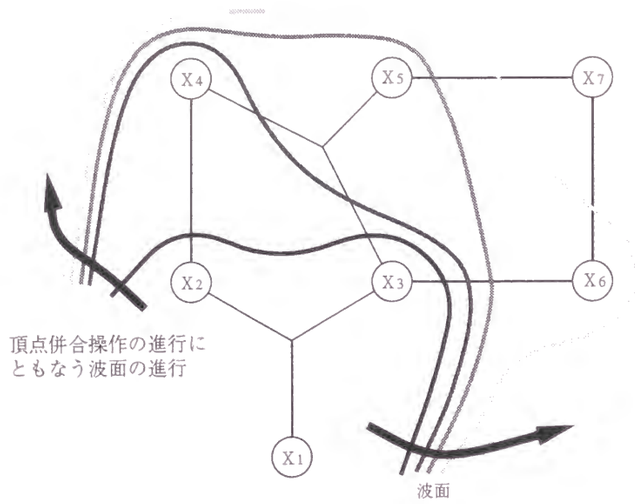


図 3.3: 併合法 (invasion) における前進過程

解構成型アルゴリズム

変数への値の割当てを，制約に矛盾しないようにしながら徐々に拡張してゆく方法で，部分解成長型の厳密アルゴリズムともいえる．このような解法としては，木探索法，弛緩法，併合法の 3 つが代表的である．

(1) 木探索法

これは通常の木探索による方法で，探索木の各レベルに変数を対応させ，分枝に値を対応させる（図 3.2）．解を一つだけ見つければよいという問題では，バックトラックを含む深さ優先探索が行われる．

(2) 弛緩法

関係する制約どうしを突き合わせて考えて無駄な要素を削除してゆく処理を弛緩あるいは制約伝播という．弛緩法の目的は最終解を求めることではなく，木探索法や併合法などの最終解探索手続きを行う前に，無効ラベルをできるだけ取り除くことによって探索空間を絞り込むための前処理と位置づけられる．

(3) 併合法

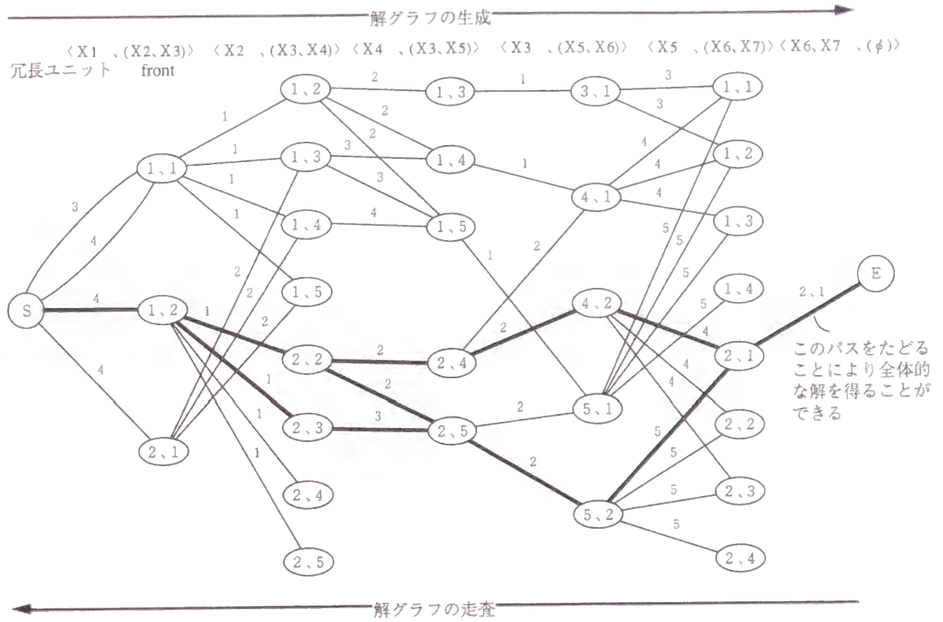


図 3.4: 併合法 (invasion) における後退過程

併合法は，複数の制約をまとめて中間解を保存しながらより多くの変数に関する一つの制約に置き換えるという併合操作を繰り返す前進過程（図 3.3）と，最終的に全変数を含む単一の制約にまで縮退した後に中間解をまとめあげ最終解を得る後退過程（図 3.4）から構成される．これは，動的計画法 (Dynamic Programming) の考え方と同じであり，全ての解を求めたい問題に適した方法といえる．

状態空間アルゴリズム

これは組合わせの試行錯誤によって探索を進めるのではなく，状態空間の地形を考慮しながら解の存在する場所への接近を試みる探索法である．例えば，状態空間のある位置，すなわち全変数に解候補としての値を仮に割当てた状態からスタートし，その制約違反の度合いが小さくなる状態に探索を進める方法が典型的である．このような近似解法としては，山登り法を用いた制約違反最少化戦略 (MCHC) や相互結合型のニューラルネットワークを用いた解法，遺伝的アルゴリズムを用いた解法などがある．

(1) 制約違反最少化戦略を用いた解法

各変数にある値が割当てられているとき，制約に違反している変数を一つ選択し，その変数に制約違反数が最少となるような値を代入することを制約違反最少化戦略という．このような制約違反最少化戦略を山登り法に適用した手法は，MCHC(Minimizing Conflict Hill Climbing)[36] と呼ばれる局所探索法である．

(2) 遺伝的アルゴリズムを用いた解法

多数の候補解の集団を状態空間内に分散させ並列的に探索を進めることを遺伝的アルゴリズムを用いて実現する手法であり，大域的な探索法といえる．これは厳密解（最適解）ではなく準最適解を比較的迅速に探索するのに有効な近似解法であるが，局所探索能力が低いために最適解を必要とする問題では，MCHC などの局所探索法との融合が必要である [37, 38].

3.2.3 制約指向の立場から見たファジィネス

例として，図 3.5に“身長 170cm 前後”というファジィ概念に対応するメンバーシップ関数を示す．通常のファジィ集合では，メンバーシップグレードは基数（数値）として設定されるが，順序関係のみが存在する序数として捉えるとき，順序付けられたクリスプな区間（制約区間）の集合体としてファジィ集合を捉えることができ，このようなファジネスの捉え方を“**区間制約ファジィ集合**”と呼ぶ [39]．上方の順序レベルは制約としてより精密かつ強い反面，制約の成立の確からしさが低い区間であるのに対し，下方の順序レベルは制約としてより粗く弱い反面，成立の確からしさが高い区間を与えている．(例えば，図 3.6に示す“山型”のファジィ集合 (a),(b) は（区間の順序集合として）同一の集合と見なされる.)

一つの区間制約ファジィ集合を一意的に表す表記法として **Min-Max 図**がある．これは各区間の下端点の値を横軸に，上端点の値を縦軸にとったデカルト座標系において，区間に対応した点の軌跡を表示したものであり，一つの区間制約ファジィ集合は一つの有向曲線（区間制約メンバーシップ曲線）で表される．

前件部，後件部がそれぞれ区間制約ファジィ集合で与えられているルールを“**区間制約ファジィルール**”と呼び，基本的につぎに二つの形式（解釈）が考えられる．

1. 可能性ルール：前件部の成立する可能性があるときは，後件部の成立する可能性がある．

2. 必然性ルール：前件部が必ず成立するときは，後件部も必ず成立する．

前者は一般に，望ましい解領域を設定する際に用いられる．後者は必ず解がその中に存在しなければならない領域を限定する際に用いられる．このようなルールは，前件部変数と後件

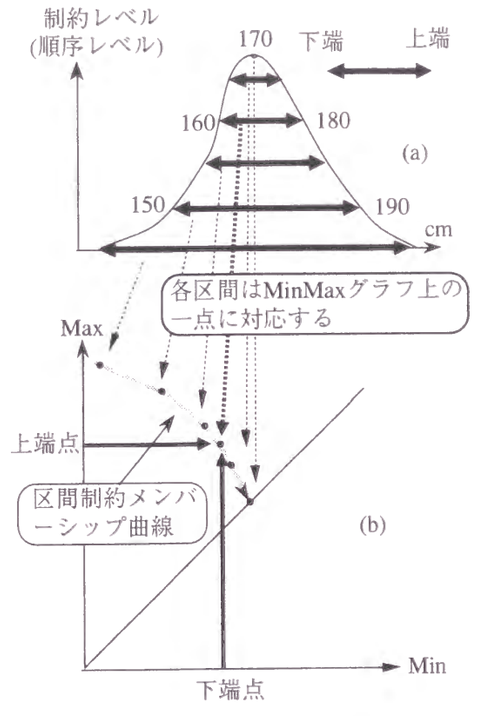


図 3.5: 区間制約ファジィ集合概念とその Min-Max 図表現

部変数に関する同時制約 C を介しての制約伝播から導かれる場合がある．この場合，図 3.7に示すように，前件部が“山型”集合のとき，後件部は“お椀型”集合となる．われわれはこのような集合も区間制約ファジィ集合とみなす．

通常のファジィ集合概念では，全ファジィ集合に共通のグローバルなメンバーシップグレード軸をもつ．これに対して区間制約ファジィ集合ではローカルな尺度としての**順序レベル（制約レベル）**軸を導入しているに過ぎない．したがって，図 3.6に示すように，複数の区間制約ファジィ集合を扱う際には，これらのローカルな順序レベル軸間を相互に対応づける“**同値レベル制約関係**”の導入が必要となる．換言すれば，処理の進行に応じて同値レベル制約関係を導入することにより，ダイナミックかつ文脈に依存した形であいまいさを取り扱うことが可能となる．このようにして，問題に内在するファジィネスをファジィ集合群からなる有機的な**ネットワークシステム**として組織化することが可能となる（図 3.6）．

このネットワークを形成する同値レベル制約関係は，**意思決定者の選好構造**を反映したものとなる．すなわち，制約レベル軸をファジィ変量の許容変動範囲（トレランス）を指定する意思決定の軸として捉えるとき，これら意思決定軸間の同値レベル性は，変量のトレラン

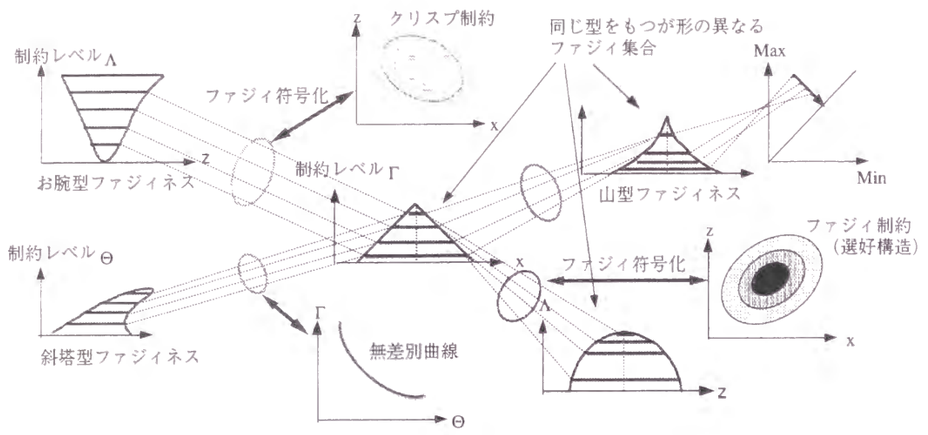


図 3.6: ファジィネスの分散ネットワークシステム

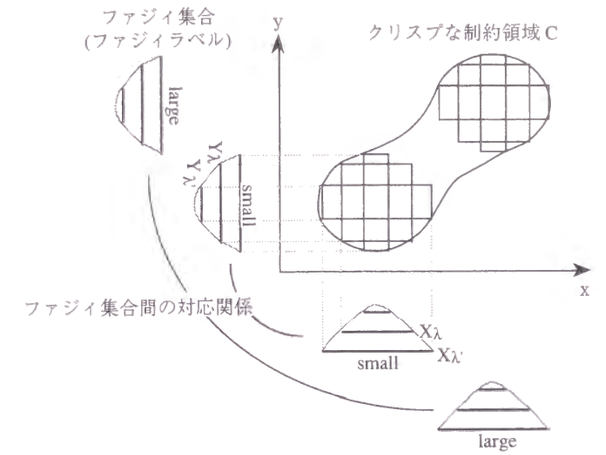
ス間のトレードオフ関係を反映したものとなる．

3.2.4 ファジィネスの導入による制約充足問題の定式化

ここでは，連続変量からなる制約充足問題に注目する．一般に制約充足問題においては，変量が多数の制約を介して互いに関係づけられており，一つの変量値の変更は，制約を介して次々と他の変量の値の変更を引き起こし，制約充足解探索の計算量の飛躍的な増大を招く．そこで，変量間に設定されている制約領域を内側から矩形（直方体）で近似することを考える．このとき，矩形をそれぞれの軸に射影した辺上のみで変量値を動かすと定めると，常に元の制約は満たされる（図 3.7）．したがって，これら制約に関連する変量を別々に扱うことが可能となる．

制約領域が複雑な場合，複数の区間制約ファジィ集合（以下，ファジィ集合と略記）を用いて，領域を近似することを考える．このとき，個々の制約区間群間の対応関係（同値レベル制約関係）と，制約区間の集合体であるファジィ集合群間の対応関係（ラベル対応関係）という二種類の対応関係を考えることになる．これらの間に適切なリンク構造を設定することによって，対象問題をファジィ集合のネットワークで表現することが可能となる．

すなわち，図 3.7に示すような，あいまいさが含まれないクリスプな制約（領域）の場合，制約 $C(x,y)$ が必ず満たされるように，制約領域を内側から複数の矩形 $X_\lambda \times Y_\lambda$ で近似する．このとき，各矩形を構成する区間 X_λ と Y_λ の間には**逆対応関係**（$X_\lambda \subseteq X_{\lambda'} \rightarrow Y_\lambda \supseteq Y_{\lambda'}$ が成立するため（図 3.7参照），“山型”のファジィ集合に対して，“お椀型”のファジィ集合（図 3.6(c) 参照）が対応づけられる．ファジィ集合間の対応関係としては，共通の矩形（直



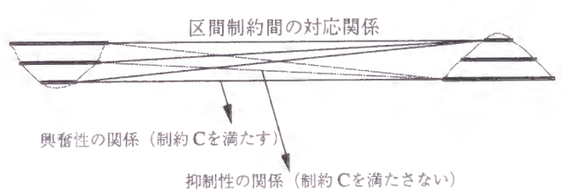


図 3.7: クリスプな制約のファジィ集合の組み合わせによる分解

方体）群構成に関与するファジィ集合群間に興奮性のリンクを設定する．また，制約区間群間の対応関係としては，同一矩形を構成する制約区間群間は興奮性のリンクで結び，そのような対応の成立しない制約区間群間には，それらが作る矩形が制約領域内に含まれれば興奮性のリンクを，一部でも制約領域外にはみ出せば抑制性のリンクを設定する．

図 3.8に示すような，多段（多レベル）の制約（領域）をここでは“**ファジィ制約**”と呼ぶ．このような制約領域群は，**制約緩和**などの**意志決定者の介在**の余地，あるいは（被制約）変量に関連する**目的関数**の達成度レベルの多段設定から導かれるものと考えられる．この場合も，制約領域の矩形による近似の考え方が，各レベル毎に適用可能となる．つまり，各レベルの制約領域について，その制約が必ず満たされるように，内側から矩形で近似する（図 3.8参照）．このとき，(一般にファジィ制約では，高いレベルの制約領域は，より低いレベルの制約領域の部分領域として与えられるために）同レベルの制約区間どうしの対応を考えると，山型のファジィ集合には山型のファジィ集合が対応づけられる．ファジィ集合群間の対応関係としては，クリスプな制約の場合と同様に，対応するラベルを表すファジィ集合間に

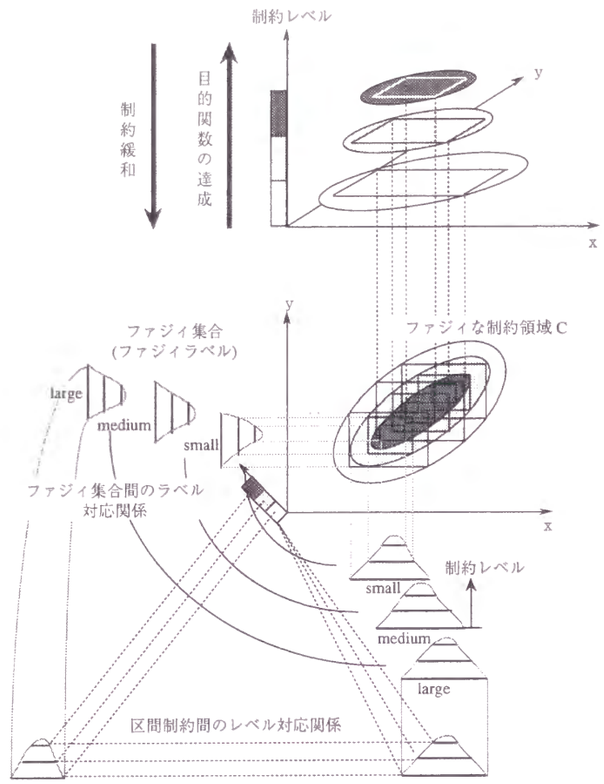


図 3.8: ファジィな制約のファジィ集合の組み合わせによる分解

興奮性のリンクを設定する．制約区間群間の対応関係としては，同一矩形を構成する制約区間群間には興奮性のリンクを設定する．

以上のように，連続変量上の制約条件に対して，その制約領域から制約区間を切り出し，制約区間の集合体としてのファジィ集合を導き，それらにラベル付けを行うことによって，連続変量を含む制約充足問題を，2.1 節で述べた整合ラベリング問題（CLP）として扱うことが可能となる．つまり，変量をユニット，ファジィ集合をラベル，ファジィ集合間の対応関係をラベル拘束関係とした CLP と捉えられる．また，制約がファジィであるかクリスプであるかにかかわらず，統一的な扱いが可能となる．この方法では，各制約領域を矩形（直方体）で近似することにより解の探索空間を有限に絞り，計算量を大幅に減少させることが可能となり，制約充足問題の近似解法を構成している．

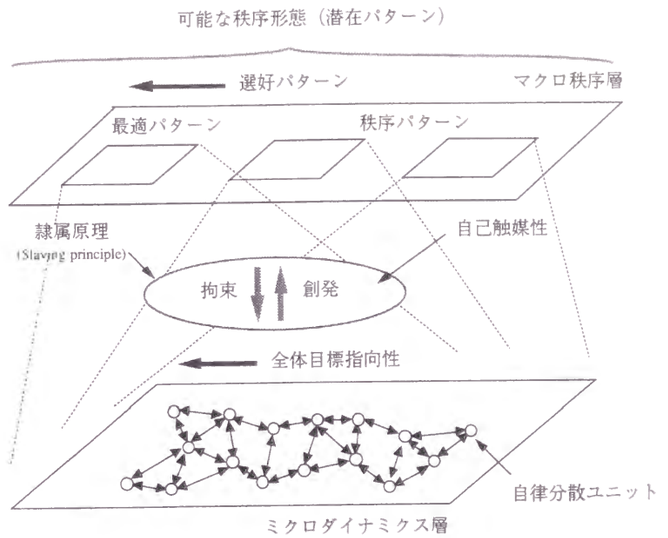


図 3.9: 自律分散型問題解決におけるシナジェティクスの原理

3.3 記号処理と自律分散型処理の融合による制約充足問題解決システム

本節では，まず自律分散システムの構成原理に関する考察について述べ，記号処理に基づく制約充足問題の解法を紹介した上で，ファジィネスを介して記号処理と自律分散型処理を融合した階層型の制約充足問題解決システムを提案する．

3.3.1 自律分散システムの構成原理

我々は既に，Haken の**シナジェティクス**（Synergetics）[10] における中心概念である**スレイビング原理**を導入し，それを成立させ秩序形成を達成するためにシステム階層間に必要となる**二重制御**とそれを支える**周縁制御の原理**を実現した自律分散システムの構成概念を提案した [41]．

前章で述べたように，シナジェティクスは幅広い領域での自己組織化の原理を追求するものであり，マクロな秩序パターンの形成とミクロな要素の振舞い（ミクロダイナミクス）の間の相互関係に基づいて秩序形成を明らかにするものであり，逆にミクロダイナミクスの集積としてのマクロな秩序の形成（秩序パラメータ）が十分大きくなると，これによってミクロな振舞いが逆に規定されるというフィードバックループによる秩序形成メカニズム－ スレイビング原理（slaving principle）－を基本原理とするものである．

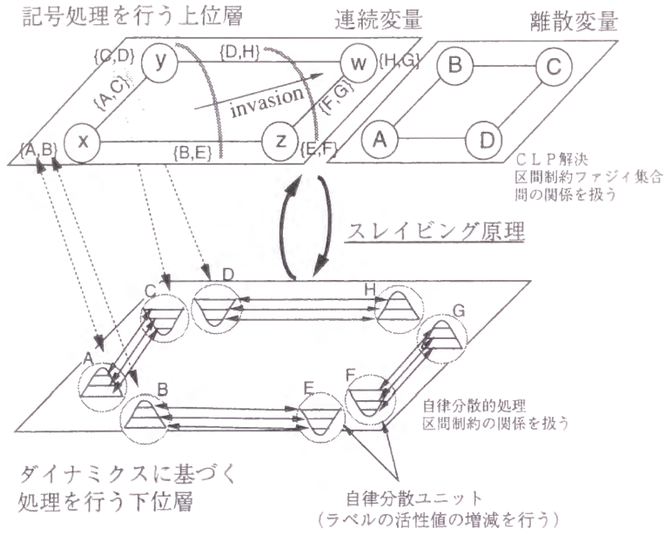


図 3.10: 記号処理と秩序形成的処理を組合わせた自律分散型問題解決システムの構成

ここでは，上位層と下位層の二層からなるシステム構成を考える（図 3.9）．下位層には，多数の“**自律分散ユニット**”が存在し，それぞれが隣接する自律分散ユニットとの間で互いに影響を及ぼし合いながら内部状態を変更し，上位層にそれを伝える．このような計算が同時並行的に繰り返されることにより上位層にはマクロな秩序が形成され，逆に形成されたマクロな秩序に下位層が隷従化する（スレイビング原理）．上位層では，2.1 節で述べた整合ラベリング問題（CLP）の解決が行われる．

スレイビング原理が成立するように，下位層の自律分散ユニットに**自己触媒性**（マクロな秩序の形成を助長するように協調的に他のユニットに振る舞いを同調させる性質）と**全体目標指向性**（形成された秩序がシステム全体に課せられた目標に沿うような性質）を付与する．これは上・下位層間の関係として，**二重制御**（各レベルは自分自身の階層を支配する法則と，上位レベルによる境界条件の決定という二重の制御を受けていること）あるいは**周縁制御の原理**（上位レベルの機能に従うように下位レベルの境界条件が決定されること）が働くように，計算場を設定することを意味する [22].

3.3.2 記号処理に基づく制約充足問題の解決

前章で述べたように，一般に CLP の解法には，木探索法，弛緩法，併合法などがある [34]. ここでは，併合法に動的計画法（Dynamic Programming）の考え方を導入して計算量・計

算時間を低減した方法である **invasion**（**侵入法**）[45] を考える．この解法は，制約ネットワーク上で front と呼ばれる波面を進行させながら，そこに含まれるユニット組に関する制約条件を**合成**（**併合**）していく前進過程と，波面進行を逆に辿りながら制約群を充足する解（ラベル組）を求める後退過程から構成される．

3.3.3 記号処理・自律分散型処理のファジィネスを介した融合による制約充足問題解決システム

2.3 節で示したように，連続変量だけでなく，**意志決定者の選好構造**，すなわち意志決定者の介在（制約緩和）・トレードオフ関係・目的関数といった要素を含む制約充足問題に対しても，ファジネスを導入して制約領域を分割し，多重（冗長）表現することによって，**さまざまな可能性を同時並行的に探索する問題解決システム**の構築が考えられる [42, 46]（図 3.10）．これは二階層からなるシステムであり，上位層における**ファジィ集合**（**ファジィラベル**）**の候補解の選択**と，下位層での（ファジィ集合中の）**制約区間の候補解の選択**が，**同時並行的かつ互いに情報交換を行いながら協調した形**で実行される．この際，先のトレードオフ関係は制約区間群間のリンク構造として，また制約緩和・目的関数は制約レベル軸上の（上位レベル優先の）**選好構造**として埋め込まれる．

下位層では，一つのファジィ集合を“**自律分散ユニット** (ADU: Autonomous Decentralized Unit)”として，それらが 2.3 節で述べた方法に従ったネットワーク構成，すなわち，同値レベル制約関係に基づいて制約区間群間に張られたリンクによって互いに関係づけられた構成がなされる．各自律分散ユニット内の制約区間は，それ自身の“**活性値**”をもっており，リンクによって結ばれた（隣接した）自律分散ユニット内の対応する制約区間との間で活性値の問い合わせを行い，それらの活性値 x_j, x_k と自らの活性値 $x_i(t)$ をもとに，次の時点の活性値 $x_i(t+1)$ を自律的に決定する．この制約区間の活性値更新においては，**自己触媒性**が実現するように，周囲の活性値が高くなれば，自らの活性値も高くなるように以下のような更新式 (1),(2) が導入されている．

$$x_i(t+1) = \frac{1}{1+e^{-u_i(t+1)}} \tag{3.1}$$

$$u_i(t+1) = \sum_j w_{ij}x_j(t) + \sum_k w_{ik}x_k(t) + x_i(t) + U_i + A_i - \theta \tag{3.2}$$

ここで，$x_i(t)$：時刻 t における区間 i の活性値，u_i：区間 i への入力，w_{ij}：異なる変数を表すファジィ集合内の区間 j との間のリンクの重み，x_j：異なる変数を表すファジィ集合内の区間 j の活性値，w_{ik}：同じ変数を表すファジィ集合内の区間 k との間のリンクの重み，x_k：同じ変数を表すファジィ集合内の区間 k の活性値，U_i：選好度に応じた初期活性値，A_i：併

合時毎の上位層からの（活性化・不活性化）入力，θ：閾値，である．このような活性値の更新が，すべての自律分散ユニットについて同時並行的に行われる．また，上でも述べたように，意志決定者の選好構造を制約レベル上の選好構造として埋め込み，上位の制約レベルがより選好され活性値が高められるように活性値更新式を設定することによって，**全体目標指向性**が実現される．

上位層では，問題を CLP と捉え，invasion によって全体的整合性を図る処理を行う．つまり，変数をユニット，ファジィ集合をラベル，ファジィ集合間の対応関係をラベル拘束関係とした CLP を解く．適当なタイミングで，下位層において活性値の高い自律分散ユニットに対応したラベル組（解の候補の可能性が高いところ）から invasion が開始される．図 3.10においては，変数 x と y との間で整合するラベル組が求められた後，front が変数 x から z に進行する例が示されている．その際，新たな front(y と z) 上で制約条件を合成する“**併合操作**”を行うことにより，許容可能なラベル組（中間解）を拡大的に求めてゆく．このような前進過程において全ての変数に関して併合が行われた後，front の進行を逆に辿りながら中間解を統合してゆく後退過程が実行され，全体的に制約を充足する解の候補が導かれる．もちろん，上位層での CLP の解法として木探索法などの他の手法をとり，下位層からの情報を上位層での処理に反映させることも考えられる．しかしここでは，**処理速度**が優れている点を考慮するとともに，全体の論理的整合性を図る記号処理（上位層）と局所的な制約充足を行う自律分散型処理（下位層）という**二つの対比的な計算原理を融合**に際し，下位層での**並行的**な処理をうまく活かした形で上位層での処理を進めることを考え，invasion を採用する．

以上のようにして構成された二つの階層間には，先に述べた**スレイビング原理**を参考にし，以下のような相互作用が導入されている（図 3.10参照）．

1. 下位層での活性値が閾値を超えた自律分散ユニット（ADU）についてのみ，その活性値（ADU 内の各区間の活性値の和 $\sum_{i \in ADU} x_i$）を上位層に伝達する．

2. invasion での前進過程における併合操作（制約条件の合成）を行う前に，下位層から活性値の伝達されなかったラベルをあらかじめ削除しておく．

3. 併合操作の結果，上位層で制約充足の可能性なしとして削除されたラベルに対し，併合時における上位層からの負の（不活性化）入力 A_i により，その下位層での活性値を減少させるとともに，削除されなかったラベルに対しては正の（活性化）入力 A_i により，下位層での活性値を増加させる．

4. 上位層での処理（invasion の前進過程）は，下位層での処理がある程度進んだ段階で開始され，下位層での処理情報（活性値）が併合操作時にうまく活用されるように，下位層での処理（活性値の更新）と比較して緩慢なペースで進行させる．

相互作用 1, 2 は，ミクロダイナミクスの集積としてマクロな秩序を形成することを目指したものであり，相互作用 3 は逆に，マクロな秩序によってミクロな振舞いが支配されることに相当する．また，下位層における自律分散ユニットは，下位層自身を支配する集合力学的な法則と，上位層での併合操作の結果による境界条件の決定（活性値の増減）という**二重制御**を受けていることになる．

上位層においては，併合操作がすべて終了すると invasion の後退過程を実行したのち，可能な候補解のすべてを意志決定者に提示する．つまり上位層では，全体的に整合するラベル（ファジィ集合）のすべての組み合わせが，解として出力される．一方，下位層においては，これら上位層の解に対応する解を全て**重ね合わせて**求まる一つの解が，制約区間群の活性値（活性値分布）として出力される．

さて，**上位層ー下位層**における計算の特徴を比較すると，**求解の多重度・冗長度：低ー高，解の合理性：高ー低，解の詳細度：低ー高，計算の進行：逐次的・集中的・離散的ー並行的・分散的・連続的，計算メカニズム：目的指向的（トップダウン）ー自己組織的（ボトムアップ）**のようになる．

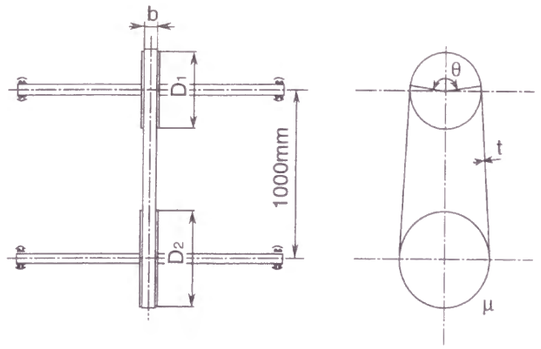
以上のようなシステムの構成および動作により，先に述べた自律分散システムの構成原理が具体化され，記号処理と自律分散型処理を融合した制約充足問題解決システムの構築が可能となる．

3.4 並列処理言語による提案システムの実装と設計問題への適用

本節では，提案システムを並列処理言語 Occam[44] によりトランスピュータ上に実装し，二種類の設計問題に対して適用を行う．Occam を用いることにより，上位層と下位層の動作の並行性，ならびに下位層における自律分散ユニット群の動作の並行性を自然な形で実装することが可能となる．

3.4.1 機械設計問題への適用

本節では，機械設計問題，とりわけ基本設計問題を取り上げる．これは，対象の全体的構造が与えられた下で，機能・寿命・信頼性・コストなど様々な要求項目のバランスをとりながら，要素諸元を絞り込んでいく過程である．この種の設計問題に現れる制約式は一般に複



与えられる変量　T ：駆動軸トルク　n ：駆動軸回転数

設計変数　D_1 ：駆動プーリ直径　D_2 ：被動プーリ直径

b ：ベルトの幅　t ：ベルトの厚さ

μ ：ベルトとプーリの摩擦係数

目的関数　σ ：ベルトの許容引張応力

W_1 ：駆動プーリの重量　W_2 ：被動プーリの重量

v ：ベルトの速度　θ ：プーリの巻掛け角

F_1 ：ベルトの張り側張力　F_e ：有効張力

A ：ベルトの断面積

i ：速比

図 3.11: ベルト伝動系の設計問題

雑なものが多く，複数の要求項目間の優先順位も明確につけにくい場合が多い．そのため，各要求項目の許容範囲をどの程度に設定するかという**意思決定者の判断**が入ることになり，あいまいさを含む制約充足問題として取り扱う必要がある．複数の要求項目のバランスをとりながら，解を求めてゆくというところに本問題解決システム適用の価値があると考えられる．設計問題にはしばしば仕様変更の伴うことがあり，設計システムの保守性・拡張性などの面からみても，自律分散的側面をもつ本システムの適用対象として，適切なものと考えられる．

具体例として，図 3.11に示すベルト伝動系の設計を取り上げる．ベルト伝動は駆動軸と被動軸に取り付けたプーリにベルトを巻き付けることによって，動力を伝達するものである．ここでは平ベルトの一段平行掛けの場合を考える．表 3.1に考慮すべき制約式を示す．ここでは駆動軸にかかるトルク T と回転数 n は与えられているものとする．設計変数は，駆動軸プーリの直径 D_1，被動軸プーリの直径 D_2，ベルト幅 b と厚さ t，ベルトとプーリ間の摩擦係数 μ である．目的関数としては，ベルトの引張り応力 σ，駆動軸プーリの重量 W_1，被

表 3.1: ベルト伝動系における制約式

$\sigma = \frac{F_1}{A}$	$F_1 = 1.5 * F_e X$
$W_1 \propto \frac{\pi D_1^2}{4} b$	$A = bt$
$W_2 \propto \frac{\pi D_2^2}{4} b$	$F_e = \frac{T}{D_1/2}$
$v = \frac{\pi D_1 n}{60 * 1000}$	$X = \frac{e^{\mu\theta}}{e^{\mu\theta} - 1}$
$\theta = \pi - 2 * \sin^{-1} \frac{D_2 - D_1}{2 * 1000}$	$i = \frac{D_1}{D_2}$

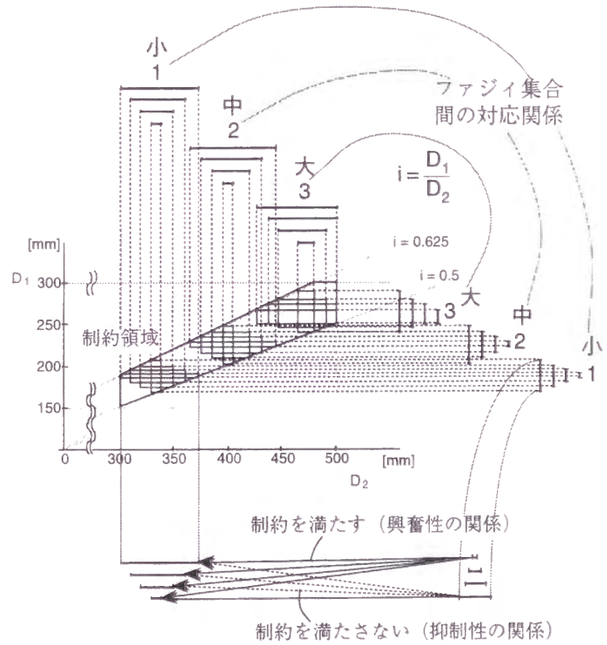


図 3.12: 制約式 $i = \frac{D_1}{D_2}$ から導かれるクリスプ制約のファジィ集合による分解

動軸プーリの重量 W_2，ベルトの速度 v，巻き掛け角 θ を考える．このとき，応力 σ を小さくすることと重量 W_1，W_2 を小さくすることはトレードオフの関係にある．

2.3 節で述べたように，各制約式中の制約領域を矩形で近似し，ファジィ集合を導く．一例として，図 3.12に制約式 $i = \frac{D_1}{D_2}$ を矩形近似し，ファジィ集合を導いた様子を示す．これ

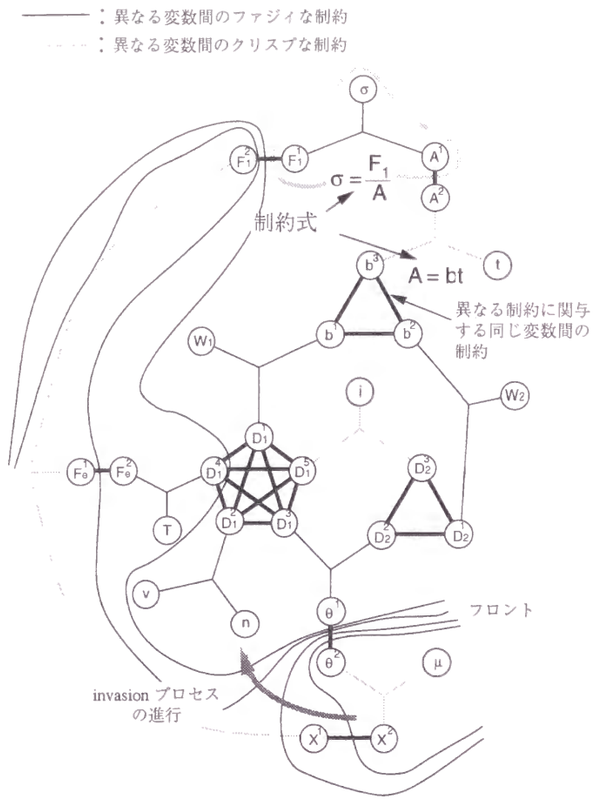


図 3.13: 記号処理層における制約ネットワークとその上での invasion プロセスの進行

らのファジィ集合が，下位層における自律分散ユニットとして扱われる．

図 3.13に，上位層で扱う制約ネットワークとその上での invasion の進行の様子を示す．また，実装システムにより得られた最終結果を図 3.14，表 3.2に示す．図 3.14は下位層での制約区間の活性状態を，表 3.2は上位層での invasion により得られた解，すなわち全体的に整合したすべてのラベル組を示している．この例題では，下位層での処理がある程度進んだ段階で，変数 μ,X,θ などに対応したラベル組の活性値が高くなり，そこから invasion が開始され，front に沿って次々と併合操作がなされてゆく．全ての変数に関して併合操作が行われた後，invasion の後退過程が実行され，表 3.2に示す 8 通りの制約充足解が得られる．

さて，最終的には，各設計変数の具体的な値を決定する必要がある．そこで，上位層で出力された候補解（ラベル組）の中から一つを選択した後，下位層でそれらに対応したファジィ集合（ラベル）を構成する制約区間の中で活性値が最大の区間を選択する．同じ変数を表す複数のファジィ集合のある場合，これら制約区間の共通部分により求めることができる．

表 3.2: 記号処理層（上位層）において選択されたファジィラベルの組合わせ

μ	X^2	X^1	F_1^2	F_e^1	T	F_e^2	$D1^4$	θ^2	θ^1	D_2^1	W_2	b^2	D_2^1	i	D_2^3	W_1	t	A^2	b^1	b^3	σ	F_1^1	A^1	D_1^3	D_1^1	v	n	D_1^5	D_1^2
1	1	2	2	2	3	3	3	1	1	1	1	1	1	2	2	1	2	2	1	2	2	2	2	1	1	2	2	2	2
2	2	2	2	2	3	3	3	2	1	1	1	1	1	2	2	1	2	2	1	2	2	2	2	1	1	2	2	2	2
1	1	2	2	2	3	3	3	1	3	3	1	1	1	2	2	1	2	2	1	2	2	2	2	3	1	2	2	2	2
2	2	2	2	2	3	3	3	2	3	3	1	1	1	2	2	1	2	2	1	2	2	2	2	3	1	2	2	2	2
1	1	2	2	2	3	3	3	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
2	2	2	2	2	3	3	3	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
1	1	2	2	2	3	3	3	1	3	3	1	1	1	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2
2	2	2	2	2	3	3	3	2	3	3	1	1	1	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2

1 つの解 (*)

ラベル： 1 ⇔ 小，2 ⇔ 中，3 ⇔ 大

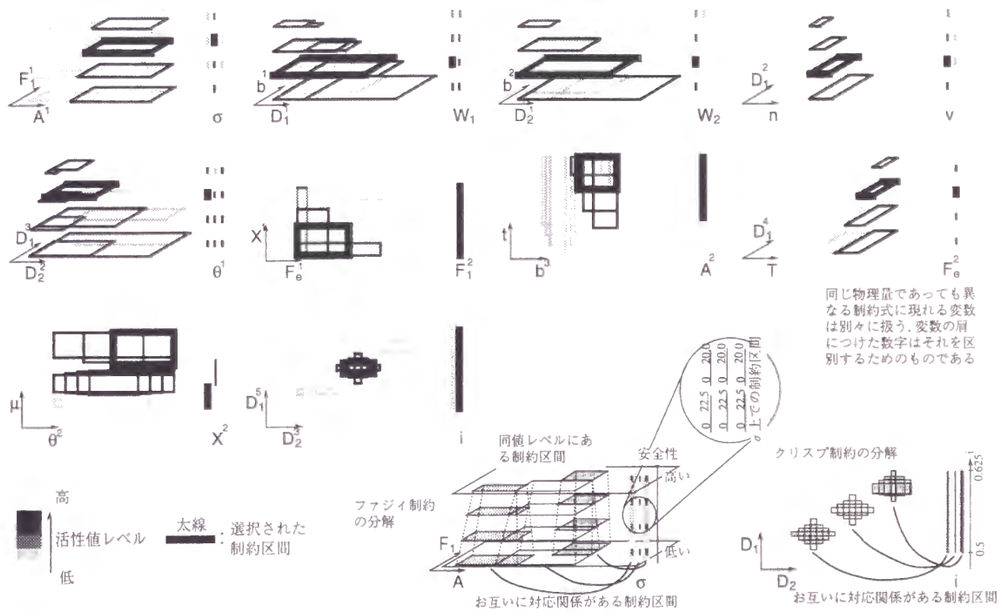


図 3.14: 自律分散型処理層（下位層）における活性値分布と選択された制約区間

表 3.2の最上段の解候補を選択した場合の最終解を表 3.3に示す．

この場合，図 3.14に示されているように，各目的関数のうち，応力 σ（図の最上段）については上から 2 段目のレベルが，重量 W_1（図の最上段）については上から 3 段目のレベルが充足されている．さて，制約式 $i = \frac{D_1}{D_2}$（図の最下段）に注目すると，D_1，D_2 は中くらいの部分が強く活性化している．D_1，D_2 が大きければ，σ は小さくなるが W_1 や W_2 は大きくなる．また，D_1，D_2 が小さければ，W_1 や W_2 は小さくなるが，σ は大きくなる．つまり，D_1，D_2 の中くらいの部分が強く活性化したということは，応力最小化と重量最小化

表 3.3: 表 3.2 の最上段の解候補 (*) を選択した場合の最終解

設計変数	許容範囲
D_1 [mm]	221.8 ～ 235.0
D_2 [mm]	312.0 ～ 360.0
b [mm]	100.0 ～ 120.25
t [mm]	5.0 ～ 6.0
m	0.37 ～ 0.50

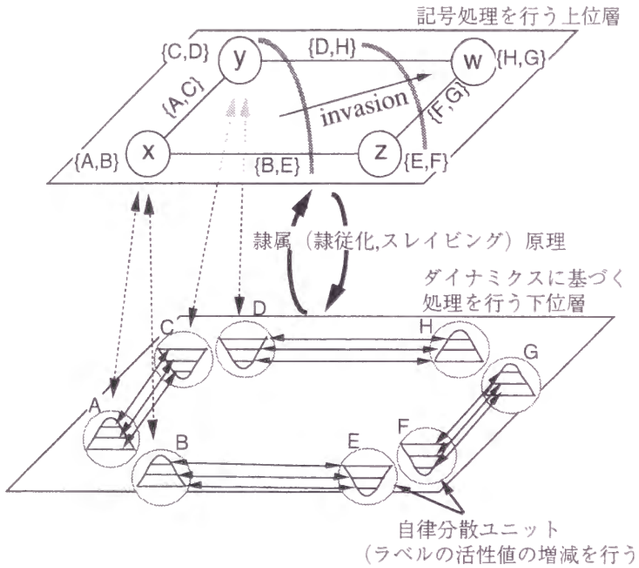


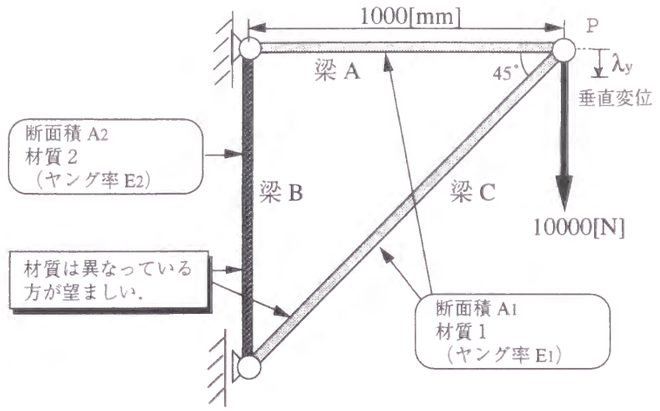
図 3.15: 記号処理と秩序形成的処理を組合わせた自律分散型問題解決システムの構成

という互いにトレードオフの関係にある要求を勘案した解が得られたことを意味している．したがって，本システムにより，複数の要求項目のバランスのとれた解を得ることができると考えられる．

上位層を外して下位層のみによる処理を行った場合，制約式 $i = \frac{D_1}{D_2}$ については，D_1，D_2 がともに小さい部分も強く活性化した．σ との関連から考えれば，D_1，D_2 は大きいほど望ましい．よって，全体的整合性を考えると上位層の導入が必要となることがわかる．

3.4.2 構造設計問題への適用 +

本節では，**連続変量と離散変量が混在**する最適構造設計問題に対する提案システムの適用を試みる．連続変量としては，部材の寸法や変位・応力などに関するものがあり，離散変量



$\lambda_y = \lambda_1 + \lambda_2$

λ_1: 梁 B による変位

λ_2: 梁 C による変位

図 3.16: 梁の設計問題

表 3.4: 梁に使用可能な材料とその特性

材質	軟鉄	鋳鉄	アルミニウム
ヤング率 E [GPa]	210	100	72
コスト係数 C [$/m3]	18	14	12
比重 r [g/cm3]	7.86	7.18	2.69

としては，材質選択や製造方法・部材の接合方法など“**構造的選択**”に関するものがある．このような構造設計問題に対して，満足解を得ることを目的に提案システムの適用を試みる．

すなわち，本章で提案する階層型の制約充足問題解決システムでは，上位層においては全体的整合性を図る記号処理が実行されるが，そこでは連続変量も記号化されて扱われるため，もともと離散値をとる離散変量も自然に扱うことができる．また，下位層では各候補値を，その“制約区間”の両端が一致して“点”になった特殊なファジィ集合として表現することにより，離散変量も扱うことが可能となる（図 3.15参照）．

図 3.16に示す簡単な梁の設計問題を取り上げる．ここで，梁 B の材質は梁 A,C の材質とは異なることが望ましいという構造的な制約が与えられているものとする．材質は，表 3.4 に示す 3 種類の金属の中から選択するものとする．また，荷重 W および梁 A の長さ L は既与のものとする．設計変数は，梁 A,C の断面積 A_1，梁 B の断面積 A_2 と，梁 A,C の材質お

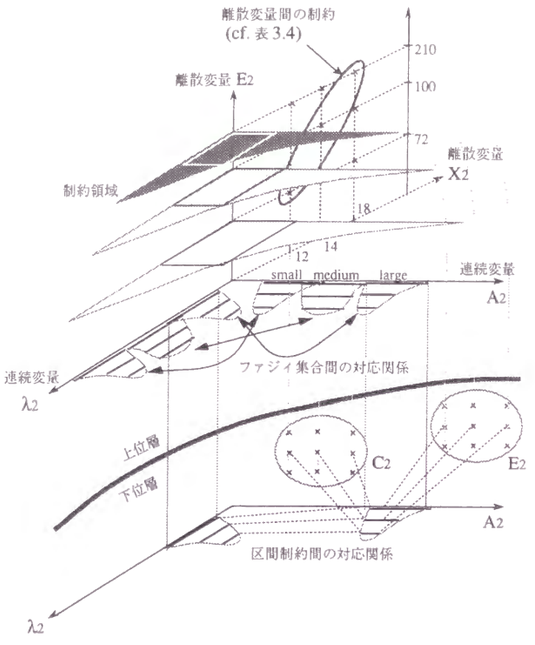


図 3.17: 連続変量 A_2, λ_2 に関する制約の分解および離散変量 E_2, C_2 に関する制約の分解から得られる上位層と下位層

および梁 B の材質を表す**離散変量**である．目的関数としては，点 P の x, y 方向の変位 λ_x,λ_y，梁にかかる応力，総重量，総コストを考える．簡単のため，コストとしては材料費のみを考え，加工や組立に要する費用は考慮しないものとする．制約式としては，これら変量からなる 7 個のものを用いた．

ここでは，先に述べたように，各制約式を矩形で近似する．矩形の選定の最適化については，ニューラルネットを応用した研究があるが，ここでは制約領域をどのような矩形で近似するのが最適かという問題は扱わないこととする．今回は，上位レベルの矩形が下位レベルの矩形に含まれるようにするということだけに注意して，制約領域を三つの矩形群で近似した．また，それぞれの矩形群は三つの区間制約より構成される．図？にそれらの矩形近似の一例を示す．(5),(6) はクリスプ制約の矩形近似を示し，それ以外ははファジィ制約の矩形近似を示している．また，図？に上位層で扱う制約ネットワークを示す．ここで，この例題においては，材料変数間には“同一の材料である”という制約が存在し，それらは上位層で取り扱われる．また 同じ物理量を表す変数であっても異なる制約式に現れるものについては，別々のユニットとして扱う．

表 3.5: 提案システムにより得られた最終解の一つ

設計変数	許容範囲
A_1 [cm^2]	21.0 ～ 24.0
A_2 [cm^2]	22.0 ～ 28.0
梁 B の材料	軟鋼
梁 A, C の材料	鋳鉄

さて，離散変量（例えば梁 B のヤング率 E_2）を含む制約式において，その候補値を設定する毎に，関連する連続変量（下方変位 λ_2 と断面積 A_2）の制約領域が決定される（図 3.17）．これは，図 3.8で示した制約レベルに応じて多段に制約領域が設定される“ファジィ制約”と同じ構造をもつものであり，これら連続変量間の関係は図 3.8と同様に扱うことができる．また，離散変量の下位層における表現に関しては，各候補値を，その“制約区間”の両端が一致して“点”になった特殊なファジィ集合として与えることにより表現でき，関連する連続変量との関係を図 3.17と同様に扱うことができる（図 3.17）．一方，離散変量どうしの関係（例えば梁 B についてのヤング率 E_2 とコスト C_2）は，図 3.17の上方に示すように，上位層における制約（CLP のラベル拘束関係）として扱うことができる．

実装システムの実行により，上位層において 11 通りの解が得られた．各変量の具体的な値を決定する際には，前節で述べた方法により値を決定する．表 3.5にこの方法により求めた最終解の一つを示す．

3.5 結言

本章では，本来あいまいさの内包されない制約充足問題に対しても，人為的にファジィネスを導入し，制約を矩形（直方体）分割することによって，**解の構造的選択**－上位層での処理－と，**解の（連続的な）許容範囲（制約区間）の選択**－下位層での処理－に還元することが可能であることを示し，**これらの選択を同時並行的かつ重畳した形で実行するシステム**として，全体の論理的整合性を図る記号処理と局所的な相互作用によって解全体のバランスをとる自律分散型処理という二つの計算原理を含む階層型の自律分散システムの枠組を提案した．この枠組による方法は，連続変量・ファジィ制約を含む制約充足問題に対する近似解法となっており，提案システムを並列処理言語 Occam によりトランスピュータ上に実装し，設計問題への適用を通して，その有効性を明らかにした．

制約のファジィネスによる**分割・還元**は，下位層での被制約変量間に相互独立性を与える

ことにより，自律分散型計算を可能にするものである．さらに，互いに重複する多数の区間群による**冗長（多重）な制約の表現**は，**意志決定者の選好構造**，すなわち意志決定者の介在（制約緩和）・トレードオフ関係・目的関数などから想定されるさまざまな可能性（候補解）を**同時並行的**に探索することを可能にしている．そこでの制約レベルとしての選好関係は，**局所的かつローカルな尺度**に基づくものであり，下位層での**並行的な**制約充足プロセスにより扱われる．また，提案システムでは，“局所的な制約緩和”が，ファジィ制約における制約レベルの緩和として，下位層での**自己組織的かつ冗長度の高い**プロセスにより実現されている．これに対し，大局的かつグローバル尺度に基づいてに定められた制約間の選好関係を，“制約階層（constraint hierarchy）”として逐次的な探索プロセスにより扱う方法もある [47]．また，制約緩和に関しては，制約の削除や拘束条件ペアの追加などにより“構造的な制約緩和”がなされた問題に対する解導出アルゴリズムが提案されている [47, 48]．我々のアプローチでは，このような制約の追加・削除などによる問題の“構造的な変更”に対しては，意志決定者による介入とその後の問題解決プロセスの再実行が必要であり，その際の介入の指針として，前回の実行時の結果を活用することが重要であり，上記のアルゴリズム [47, 48] を参考にすることが考えられる．

本章で提案したシステムは，**自律分散システムの構成原理**に基づいて，記号処理を行う上位層と連関した形で進行する計算場（下位層）での自己組織化機能に秩序形成（問題解決）を委ねるものであり，**自律的秩序形成**のために**シナジェティクス**の基本原理であるスレイビング原理が発現するように，上位層・下位層間に相互作用を設定した．意志決定者は，システムの実行により得られた解に対して（満足解かどうか）総合的な判断を下すとともに，システム内の種々のパラメータ（初期活性値，リンクの重み，invasion のタイミング等）の設定を行い，意志決定のレベルを一段上げる形で問題解決に関わることになる（**周縁制御の実現**）．もし，満足解でないと判断した場合には，上述のように，意志決定者による介入（システムパラメータの変更）により計算場を再設定した後，問題解決プロセスを再実行することが考えられる．このようなメタレベルにおける意志決定のフィードバックを含んだ問題解決のシステム化は，今後の検討課題の一つである．また，提案システムでは起こりにくいと考えられるが，上位層での併合操作により解の可能性が高いラベルまで削除される場合への対応も，今後の検討課題として挙げられる．

第 4章

制約指向問題解決プロセスに潜む複雑性

4.1 緒言

制約指向の観点からファジィネスについての新たな解釈・意味づけを与えることにより，知的情報処理，知的制御のための新たな自己組織化・学習・適応機能を有する問題解決の枠組みを導入するとともに，これらに基づいた自律分散問題解決システム [43]，ファジィ制御システム [40] などの構築についての提案がすでになされている．

本章では，これらのアプローチの根底にある，ファジィネスを介した連続変量の記号化（符号化）と，それに基づく制約充足問題解決システムの挙動について，記号力学系の立場から解析を加え，そこに内在するカオス構造の究明について検討する．さらに，システムに内在する階層構造，自己組織化メカニズムと関連づけて，カオス構造の存在意義について検討を加える．

前章で述べたように，制約指向の観点からファジィネスを捉えるとき，一つのファジィ集合は区間（制約区間）の集合組織体として解釈することができ，これを区間制約ファジィ集合と呼ぶ．このようなファジィ集合を介することにより，連続変量をファジィ集合（ファジィラベル）として記号化（符号化）することができる．このようなファジィネスによる記号化の特色としては，変量を単一の区間で区切る通常の記号化に比べ，区間の束であるファジィラベルの選択とその上での制約レベルの設定という意思決定の余地が残されているだけ，柔軟性のあるアプローチといえる．

記号化の媒体としてのファジィネスの要因としては，以下のものが考えられる．

(1) 自然言語におけるあいまいな概念にまつわるもの

(2) 専門家が経験的に獲得した操作変量上のファジィネス

(3) 問題固有の制約から自然に導かれるファジィネス

(3-1) クリスプな同時制約から導かれるファジィネス

(3-2) ファジィな同時制約から導かれるファジィネス

ここにファジィな制約とは，変量間に課せられる制約が一意的に決まるのではなく，緩やかなもの（下位制約レベル）から狭く厳しいもの（上位レベル）まで多段に設定され，問題解決主体により何れを選択するかの意思決定の余地が残されているものを指す．(3) では，このような同時制約（クリスプな一段レベルのものも含めて）を，制約構成要素の各変量に射影する形で制約を分解することによって，自然に変量上にファジィネスが導かれると考えるものである．

すなわち，前章で述べたように，クリスプな同時制約の場合（図 3.7参照），制約領域を内側から矩形で近似することにより，同時制約を同じ矩形を構成する区間どうしの対応関係と各区間についての区間制約の二つに分解する．また，ファジィな同時制約の場合（図 3.8参照），クリスプな同時制約がグレード軸方向に重なって多段の制約が構成されているものとみなすことにより，制約領域の矩形による近似の考え方が，そのまま各レベル毎に適用可能となる．したがって，それらの矩形を x 軸，y 軸に射影すると，各変量上に区間制約ファジィ集合が得られる．このように，連続変量上の制約条件に対して，その制約領域から制約区間を切り出し，制約区間の集合体としての区間制約ファジィ集合を考え，これをラベルとして扱うことによって，連続変量を含む制約充足問題としての取り扱いが可能となる．

以下，まず **4.2節**において，制約伝播による問題解決を考え，制約伝播の際に記号化を介さない “制約伝播力学系” の振る舞いについて数学的に調べるとともに，計算機シミュレーションにより確認を行なう．また，**4.3節**においては，ファジィネスによる連続量の記号化（符号化）を導入した “ファジィ記号力学系” に内在する複雑性を計算機シミュレーションを通して明らかにする．

4.2 制約伝播力学系

4.2.1 制約伝播と問題解決

制約充足問題の一つの解法として，局所的に制約を充足させることを繰り返し積み重ねていくことで全体の制約充足をおこなう方法があり，これを制約伝播法という．

例として，図 4.1に示すように，変数 x,y,z があり，x と y，y と z，z と x の間には，それぞれ固有の制約 A,B,C があるとする．仮に，x に何らかの初期値を与えると，制約 A により y が決定される．y が決定されると，つぎは制約 B によって，z が決定され，さらに，制約 C によって再び x が決定されるといったように局所制約充足を繰り返していく．x,y,z の各変数が一定値に収束したならば，すべての制約を充足する解が求められたことになる．

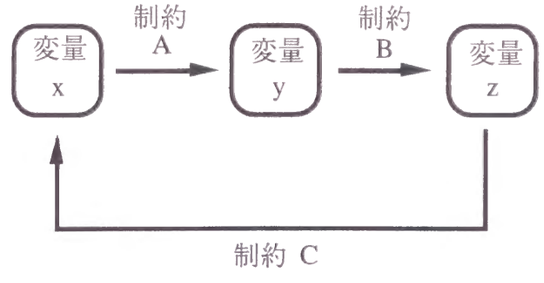


図 4.1: 制約伝播の流れ

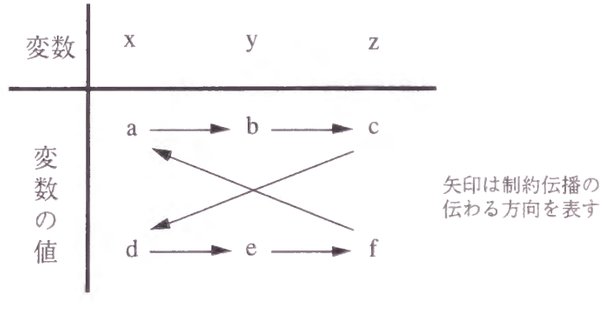


図 4.2: 制約伝播における 2 周期解

しかし，制約伝播を繰り返しても収束するとは限らず，図 4.2に示すように，2 周期解，さらには多周期解になることもありうる．そこで，この周期解の意味について考えてみる．各瞬間において 3 変数はそれぞれある値をとっているが，このとき 3 制約のうち必ず 1 つは満たされていないことがわかる．すなわち，矛盾がどこかで生じており，つぎの瞬間にはその矛盾は解消されるかわりに他の矛盾が生じている．常にシステムは目先の矛盾を解消することで，全体のバランスはとられているのである．これは社会システムのような動的 (dynamic) なシステムではしばしば見うけられ，一つの問題解決の形態とみなせる場合もあるが，設計のような静的 (static) なシステムに対しては完全な解を与えるものではない．

4.2.2 制約伝播力学系とその安定性

本節では，制約を介して，区間（制約区間）を順次伝達するような制約伝播を考える．例

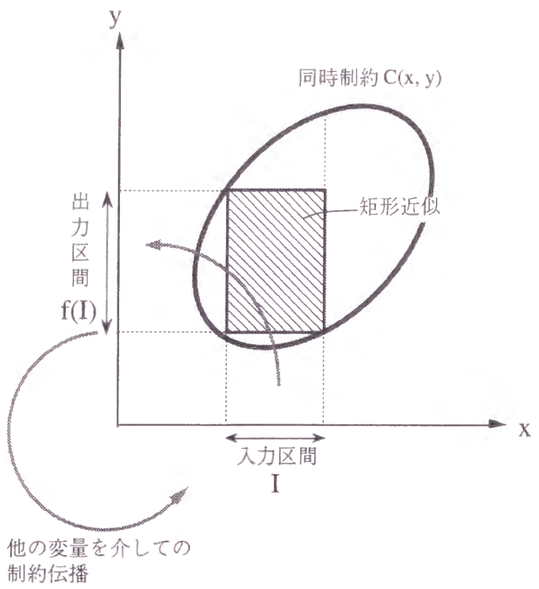


図 4.3: 制約伝播力学系

えば，図 4.3に示すように，変数 x, y の間の制約として同時制約領域 $C(x, y)$ を考え，変数の値ではなく区間（制約区間）が逐次的に伝播されるとする．このとき，制約領域を内側から矩形で近似することにより制約伝播後の区間が得られると考える．制約伝播後の出力区間は，制約領域から決まる関数と入力区間から求められるので，このような制約伝播プロセスは一種の力学系を構成すると捉えることができる．これを**制約伝播力学系**と呼ぶことにする．

以下では，制約伝播の伝播軌道が収束すること，すなわち，制約伝播力学系の安定性について述べる．まず，制約伝播のもつ，以下の“単調性”に注目する．

$$I' \leq I \ \Rightarrow f(I') \geq f(I) \tag{4.1}$$

I, I' は制約伝播される制約区間であり，$f(I'), f(I)$ は伝播された区間を表す．また，“$\leq$”は区間の間の包含関係を指す．

式 (4.1) は，以下のように示される．まず，図 4.3に示すように，2 つの変量 x, y の間に同時制約 $C(x, y)$ があるような制約伝播系について，2 つの入力区間 I, I' $(I' \leq I)$ があるとする．これらに対応する出力区間をそれぞれ $f(I), f(I')$ とする．すると，$I, f(I)$ は次式を満たす．

$$(\forall x \in I)(\forall y \in f(I)) \ \ C(x, y) \tag{4.2}$$

すなわち，x の入力区間 I の下で，$C(x, y)$ を満たす y の出力区間のうちで最大のものが

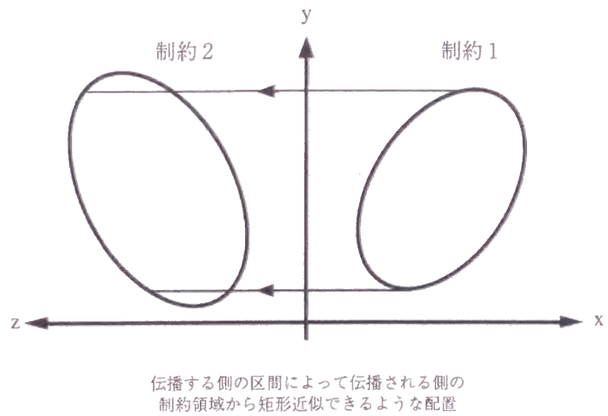


図 4.4: 制約領域の配置前提

$f(I)$ となる．同様に，$I', f(I')$ も次式を満たす．

$$(\forall x \in I')(\forall y \in f(I')) \ \ C(x, y) \tag{4.3}$$

ここで，$I' \leq I$ と式 (4.2) より次式が導出される．

$$(\forall x \in I')(\forall y \in f(I)) \ \ C(x, y) \tag{4.4}$$

この式 (4.3) における $f(I')$ は $C(x, y)$ を満たす y の出力区間の最大のものであり，したがって，式 (4.3) , (4.4) より，

$$f(I') \geq f(I) \tag{4.5}$$

を得る．

したがって，I から $f(I)$ への一回の伝播では，このように逆単調関係になり，もう一回伝播すれば順単調関係になる．つまり，制約伝播のサイクル内にある同時制約の個数が奇数か偶数かによって，逆単調関係か順単調関係が決定される．ここでまず，つぎの 2 つの定義を導入する．

定義 1 （永劫回帰領域）

その区間から出発し，上記の制約伝播を無限回続けることができるような初期区間の集まりを永劫回帰領域 D と呼ぶ

定義 2 （中間性関係）

$$bet(I : I', I'') \equiv I' \leq I \leq I'' \ \ or \ \ I'' \leq I \leq I' \tag{4.6}$$

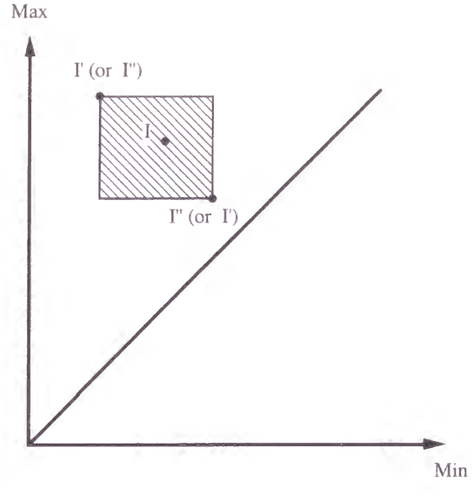


図 4.5: 中間性関係

定義 1 の前提としてあるのは，図 4.4に示すように制約伝播される側は，制約伝播する側のどの領域から矩形（区間）が伝播されてきても矩形が切り出せるように，伝播する側の制約領域を受けとめるように配置される必要がある．すなわち，入力区間が制約領域よりも大きくならないようにする必要がある．また，定義 2 は，Min-Max 図上では，図 4.5に示すように I' と I'' が相互に右下がり（左上がり）の位置関係にあるときに，I がこれら I' と I'' で構成される斜線部領域に含まれることを意味している．この 2 つの定義と先の単調性により，以下の 3 つの定理を導くことができる．

定理 1

D はつぎの意味で凸である．$I', I'' \in D$ ならば，$bet(I : I', I'')$ なる任意の I について $I \in D$ が成り立つ

定理 2

D は単連結である

定理 3

$\exists I \in D\ s.t.\ f^*(I) \leq I\ or\ f^*(I) \geq I$ ならば D は不動点をもつ

ここで $f^*(I)$ は I が制約伝播サイクルを通ってもとの変量の制約区間に戻ってきたときの区間を表す

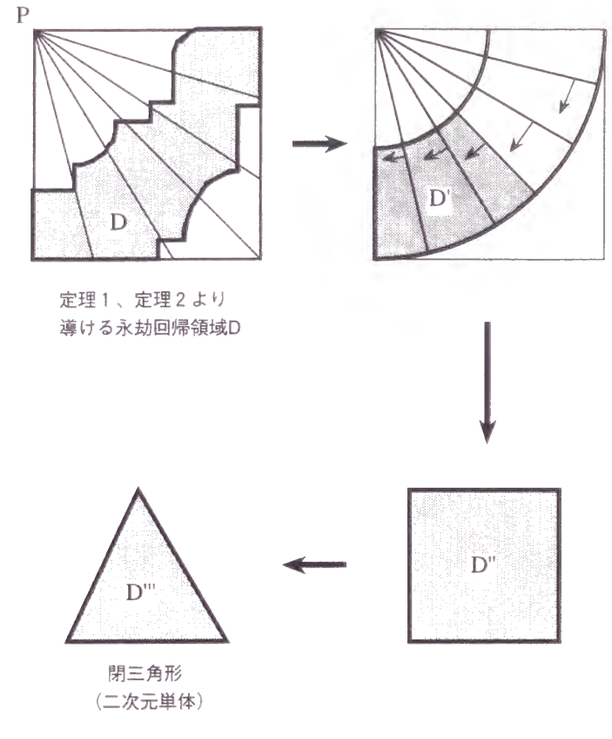


図 4.6: 永劫回帰領域 D と同相な図形

上記の定理より，制約伝播力学系には，安定か不安定かはわからないが，不動点が存在することがわかり，伝播軌道が単周期軌道に収束する可能性があることを意味する．

一般に不動点の存在に関しては，以下に示す *Brouwer* の不動点定理がある．

***Brouwer* の不動点定理**

n 次元空間 X からそれ自身への任意の写像 f について X が n 次元単体 $|\sigma_n|$ に同相であれば，f は少なくとも一つの（ $f(x) = x$ ）となる不動点 $x\ (\in X)$ をもつ．

今まで示してきたような矩形の伝播系では，空間 X は二次元であり，二次元単体は閉三角形領域（二次元単体）を意味する．上記の定理 1 および定理 2 だけからは，永劫回帰領域 D は右上がりの曲線と水平，垂直の直線で囲まれる閉領域となることはわかる．ここで，永劫回帰領域 D にくびれがないと仮定すると，図 4.6に示すように永劫回帰領域 D と 2 次元単体である閉三角形領域が同相であることが示せる．

ここで，図 4.7に示すような 4 つの同様の楕円型制約領域をもつ単純な制約伝播力学系について考える．矩形の切り出し基準点となる端点 A_1, A_2 に注目することによって，安定不

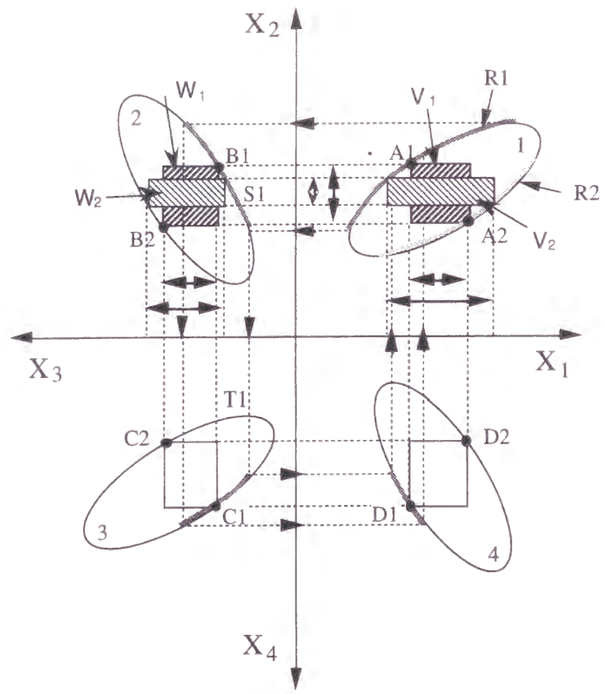


図 4.7: 端点に注目した不動点定理による不動点の証明

動点の存在を示す．矩形の伝播が周期軌道に収束している状態は，端点 A_1, A_2 の推移がそれぞれ周期軌道に収束している状態と考えられる．つまり，4 つの制約間で伝播が一周行われることを写像 f^* と表現すると，端点 A_1, A_2 は不動点ゆえに $f^*(A_1) = A_1$, $f^*(A_2) = A_2$ と表せる．これは，*Brouwer* の定理を用いて証明できる．楕円 1 の境界上で端点 A_1 をとることが可能な領域は図 27 の R_1 で示される区間である．A_1 は楕円 2 の境界上の点 B_1 に伝播される．定義 1 の前提より，制約 2 は制約 1 を受けるように配置されているので，R_1 の区間全体が伝播されると考えると，伝播されたほうの区間 S_1 は，長さが $S_1 \leq R_1$ となる領域に収束する．このように伝播が 4 つの制約間で一周繰り返されると，R_1 の領域は R_1 の内部に写像されることがわかる．この場合，*Brouwer* の定理における空間 X には，曲線 R_1 が相当する．曲線は一次元なので，その単体は線分（一次元単体）になる．曲線 R_1 を線分 Y_1 に変換する同相写像を k とすると $Y_1 = kR_1$ となるので，Y_1 と R_1 は同相として扱うことができる．したがって，R_1 が R_1 内部に写像されることが示され，さらに連続であることは自明なので，*Brouwer* の定理より R_1 の領域中には少なくとも 1 つ不動点が存在するといえる．もう一方の端点 A_2 についても，楕円 1 の境界上で端点 A_2 を取ること

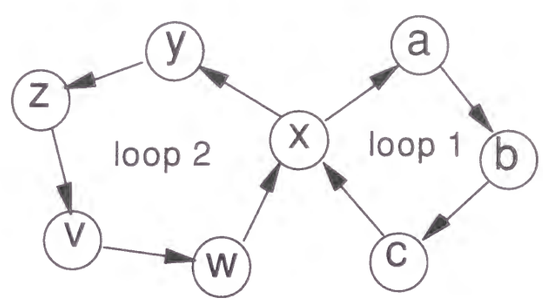


図 4.8: 2 つのループから構成される制約ネットワーク

が可能な領域である R_2 は，同じく R_2 に写像される．よって，R_2 も同様に R_2 内部に少なくとも 1 つの不動点が存在するといえる．2 つの端点がそれぞれこれらの不動点に収束することにより，矩形による伝播が周期軌道に収束することがわかる．

4.2.3 パレート最適解の導出

制約領域を矩形で近似するときの制約伝播では，前章 3.2.3 節で述べた必然性ルールを用いる．これは，入力が制約を常に満たしていることを保証しつつ，出力範囲を最大限に取るためである．このとき，入力区間と出力区間の両方がともにより広い区間となる矩形で切り出すほうが，精度よく制約領域を近似するとともに，変量としてのトレランス（許容範囲）も広いために望ましい．矩形の伝播が周期軌道に収束したときは，すべての制約について，それ以上に大きな矩形で切り出せるものが存在しないことという意味での最適性を有している．

例えば，先の図 4.7で示したように，4 つのクリスプな制約の間で伝播を行い，その伝播が周期軌道に収束したときを考えると，制約 1 の入力の区間をより大きくしようと考えて，矩形 V_1 を矩形 V_2 に変化させて，伝播を再開させると，制約 1 から出力される区間は，制約伝播のもつ先の（逆）単調性より，得られた V_2 のほうが元の V_1 よりも必然的に狭くなる．したがって伝播された側では，制約 2 で切り出されていた矩形 W_1 は，それよりも入力区間の狭い矩形 W_2 に変化する．1 つの制約の入力区間を大きくすることが，他の制約の入力区間を小さくしてしまうわけである．

このように，「他の効用を低下させることなしには，ある効用を高め得ない状態で，資源配分が最も効率的に行われている状態」は，一般にパレート最適解という．したがって，制約伝播が周期軌道に収束した場合は，各変量上に得られた制約（制約区間）の組は，パレー

ト最適な解となっていることを意味する．

前節で述べたように，制約伝播力学系は安定的な性質を有するために，上で述べた制約伝播を繰り返すことでパレート最適解を導くことができる．

4.2.4 制約伝播力学系の計算機シミュレーション

シミュレーションモデル

本節では，制約伝播力学系に内在する性質・構造を調べるために，以下に示すように系を単純化したうえで計算機シミュレーションを行なう．

1. 変量は最大 4 変量までとする．
2. 制約は単方向のループ構造を形成し，2 変量間の制約のみを扱う．
3. 制約領域の形状は，クリスプな楕円制約と，入力＝出力となる傾き 45 度の直線制約のみを扱う．
4. 各変量では過去（前回）にとった値をある一定の割合だけ考慮する．すなわち，変量 x が x_n という値（区間）をとり，制約伝播のループを通して再び戻ってきたときの値を $f(x_n)$ とすると，新たな値 x_{n+1} は，α を記憶係数（学習率）として，以下のように与えられる．

$$x_{n+1} = \alpha x_n + (1-\alpha) f(x_n) \tag{4.7}$$

5. 2 つのループをもつ制約ネットワークの場合（図 4.8），共通変量 x に異なる値（区間）x_l, x_m が同時に入ってきたときの入力区間は，妥協値（区間）としてそれらの平均値をとる．
6. 区間が楕円制約の幅よりも大きくなった場合は，その区間を楕円の幅まで縮小させる．

シミュレーションでは，単一ループの制約ネットワークから構成される系と 2 つのループをもつ制約ネットワークから構成される系について，その振る舞いを調べる．

シミュレーション結果と考察

(1) 単一ループの制約ネットワークの場合

図 4.9に示すように，4 つの変量間についてクリスプな楕円制約を設定する．楕円の傾き $\theta_2 = 50$ 度，$\theta_3 = 20$ 度，$\theta_4 = 45$ 度として，θ_1 を 10 度から 80 度まで変化させて制約伝播を

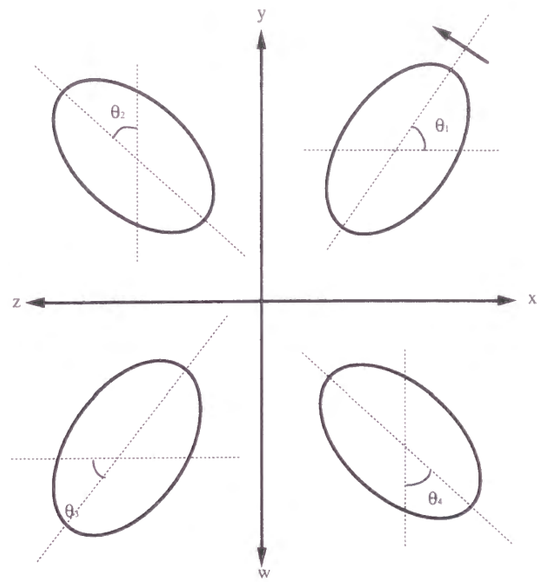


図 4.9: ループが 1 つの場合のシミュレーションモデル

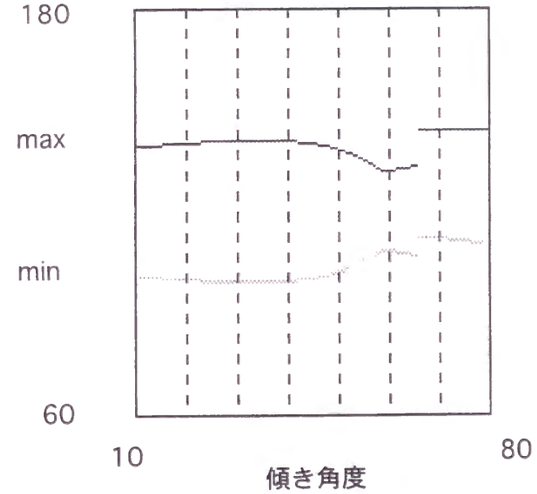


図 4.10: 単一ループから構成される制約伝播力学系の分岐図

行なったときの変量 x 上の区間に関する分岐図を図 4.10に示す．分岐図は，横軸に傾きの角度 θ_1，縦軸に区間の値をとり，max が濃色で min が淡色で表示している．また，$\theta_1 = 30$ 度としたときの制約伝播の様子を図 4.11に示す．

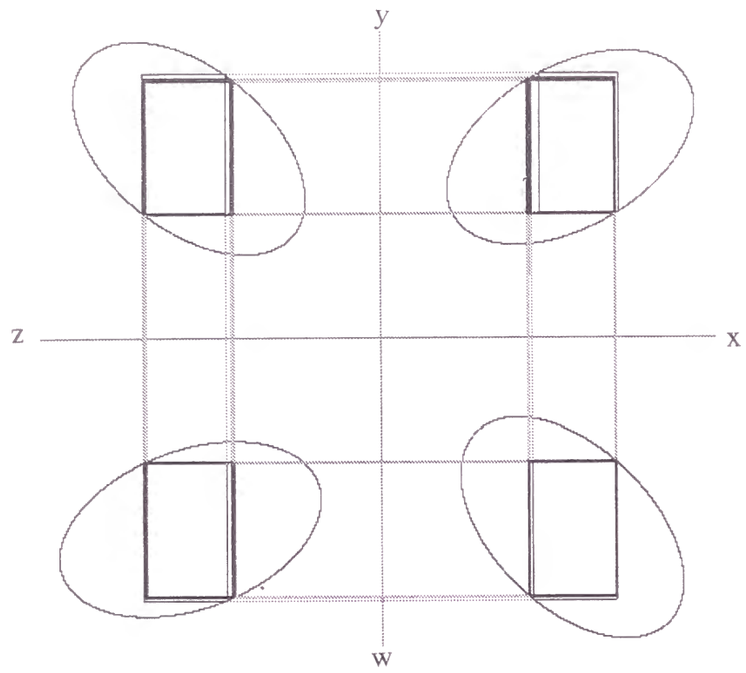


図 4.11: 制約伝播力学系における制約伝播の様子

これらの図からわかるように，どのような θ_1 の値に対しても，この系は 1 周期解に収束している．

(2) 2 つのループをもつ制約ネットワークの場合

図 4.12に示すように，周期の異なる 2 つのループ（4 周期と 3 周期）を設定する．楕円の傾き $\theta_1 = 60$ 度，$\theta_3 = 20$ 度，$\theta_4 = 50$ 度，$\theta_5 = 30$ 度として，θ_2 を 10 度から 80 度まで変化させて制約伝播を行なったときの共通変量 x 上の区間に関する分岐図を図 4.13に示す．この図より，ある角度以上になると，分岐が起きていることがわかり，$\theta_2 = 60$ 度では 6 周期解に，$\theta_2 = 70$ 度では 12 周期解に収束している．

これらのシミュレーション結果より，2 つのループをもつ系においても，楕円制約の数や傾き角度，ループの周期などによって周期は異なるが，周期解に収束することがわかり，単一ループの場合と同様に，系の安定性が示された．これらの結果より，さらに多くの変量からなる複雑なネットワーク構造をもつ系も，上記の性質をもつループを組合わせたものと考えられ，周期解に収束する安定な振る舞いをすることが予想される．

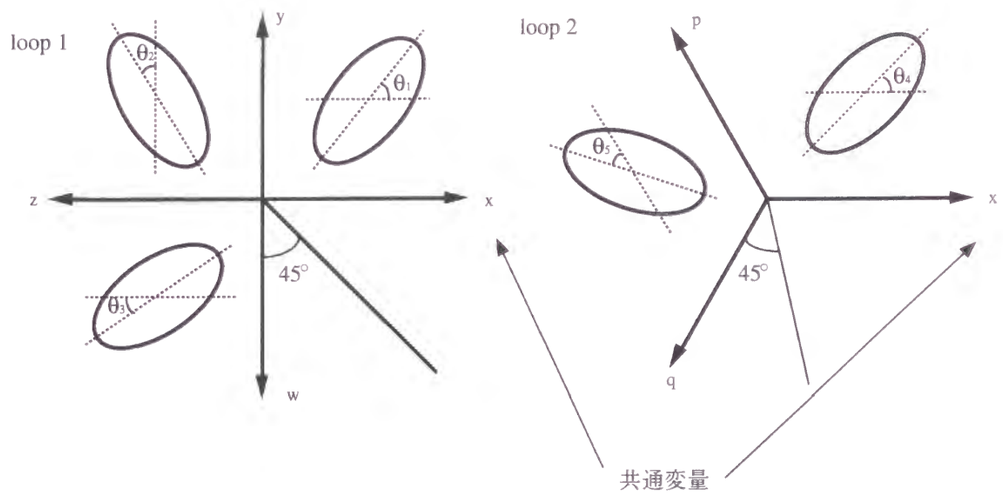


図 4.12: ループが 2 つの場合のシミュレーションモデル

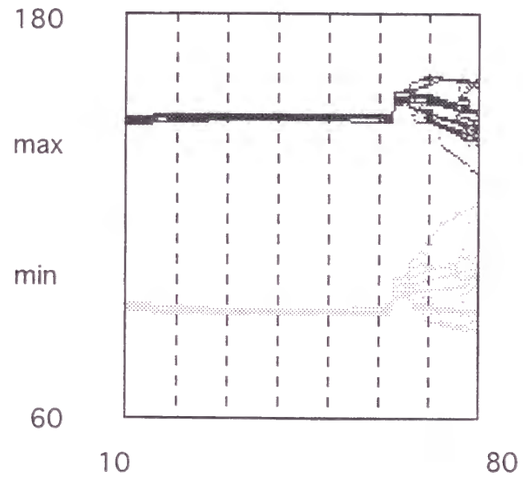


図 4.13: 複数のループから構成される制約伝播力学系の分岐図

4.3 ファジィ記号力学系と内在する複雑性

4.3.1 ファジィ記号力学系の導入

ここでは，上に述べた制約伝播力学系において，変量に固有なファジィネス（区間制約ファジィ集合）を導入し，そのファジィネスを介した“記号化”（ファジィラベルの選択）と“整形”（制約レベルの選択）を伴う制約伝播を考える．

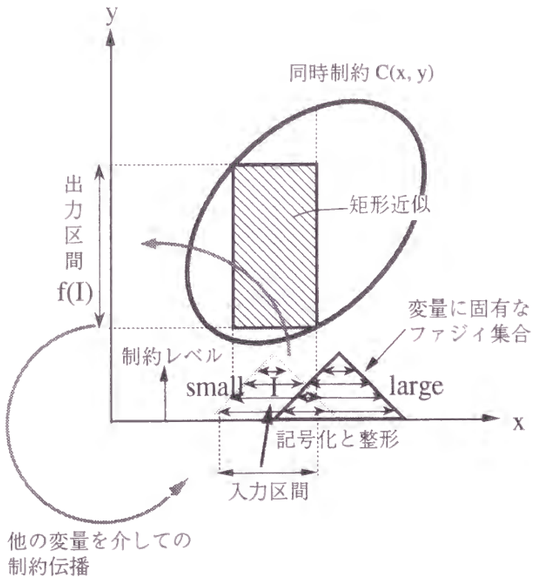


図 4.14: ファジィ記号力学系

すなわち，変量に固有なファジィネスとして，複数のファジィ集合（ファジィラベル）が設定されているものとする．この変量に対してある入力区間が与えられたとき，一定のルールにより，いずれかのファジィ集合が選択される（ファジィラベルの選択）とともにそのファジィ集合に属するある区間が選択され（制約レベルの選択），その選択された区間によって制約領域を矩形近似する．この矩形近似により得られた区間は，つぎの変量に対する入力区間となり，次々と制約伝播がおこなわれる（図 4.14参照）．

このような制約伝播力学系において，ある変量に関してどのファジィラベルが活性化されたか，すなわち，“記号（ファジィラベル）”によってこの力学系の構造を捉えようとすることは，記号力学の考え方に沿うものである [53]．そこで，このような制約伝播力学系を**ファジィ記号力学系** (Fuzzy Symbolic Dynamics) と呼ぶことにする．

前節で述べたように，ファジィネスを介した“記号化”と“整形”を伴わない制約伝播力学系は，ある意味で単純な振る舞いをし，安定平衡点へ収束する安定的な性質をもつが，ファジィネスを介して，“記号化”（ファジィラベルの選択）と“整形”（制約レベルの選択）のメカニズムを導入すると，ある種の“ゆらぎ”が組み込まれ，複雑な振る舞いをすることが推測される．

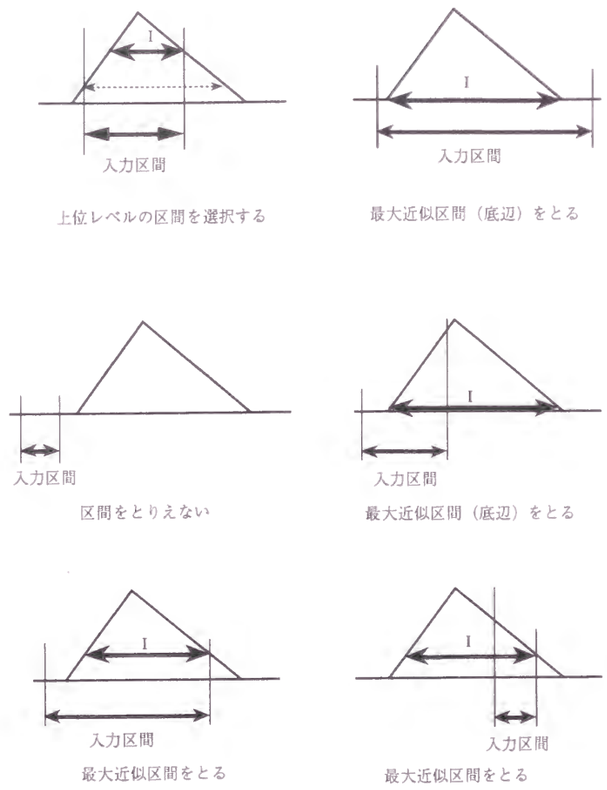


図 4.15: 入力区間から候補区間を決定するルール

4.3.2 ファジィ記号力学系におけるカオス的挙動の生成

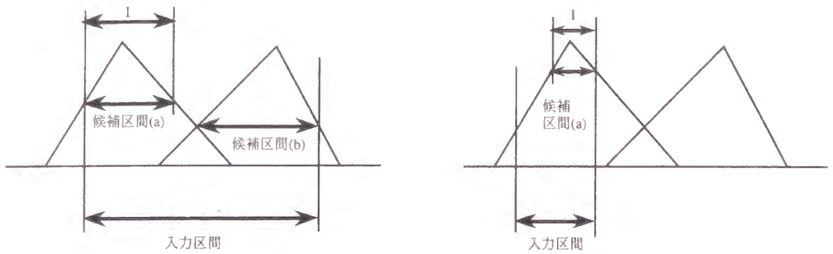
そこで，ファジィ記号力学系に内在する基本的な性質・構造を調べるために，(1) 図 4.14 に示されるような単一ループの制約ネットワークから構成される系や (2) 図 4.8に示されるような複数のループをもつ制約ネットワークから構成される系について，計算機上で制約伝播シミュレーションをおこなった．

シミュレーションモデル

本節では，前節で述べたファジィ記号力学系に内在する性質・構造を調べるために，以下に示すように系を単純化したうえで計算機シミュレーションを行なう．

1. 基本的には 4.2.4 節で述べたシミュレーションモデルに対して，ファジィネスを導入したものである．

2. 各変量に固有なファジィ集合はそれぞれ 2 つとする．これらを用いて区間を選択する



候補区間が複数あるとき
上位レベルの区間を選択

候補区間がただ 1 つのとき
その区間を選択

図 4.16: 候補区間から新しい区間を決定するルール

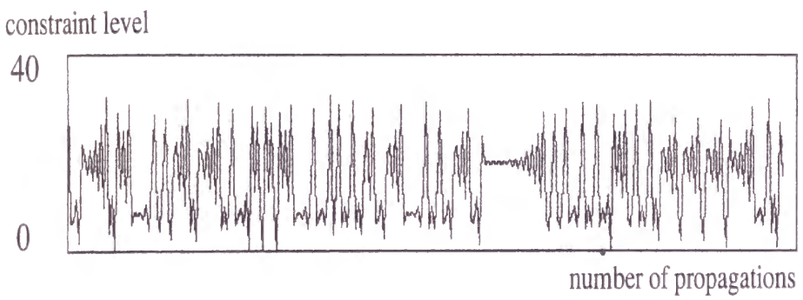


図 4.17: 選択される制約レベルの遷移

ルールとしては，2 種類のものを考える．各変量に対する入力区間より，各ファジィ集合における候補区間を定めるルール（図 4.15）と，各候補区間よりいずれのファジィラベルの区間を選択するかを決定するルール（図 4.16）である．

シミュレーションでは，初期入力区間を与える変量 x において，選択される区間（制約レベル）の変動を調べるほかに，ファジィラベルの選択を記号列で表し，その推移を観察する．

ここでは，最も単純な系である楕円制約が 1 つで単一ループの場合のほかにも，楕円制約が複数の場合や 2 つのループをもつ制約ネットワークの場合についても調べる．

シミュレーション結果と考察

[制約レベルの解析]

図 4.4 に示すようなファジィ記号力学系において，変量 x に対してある初期入力区間を与え，制約伝播を行った際の制約レベルの推移例（時系列）を図 4.17に示す．この時系列に関

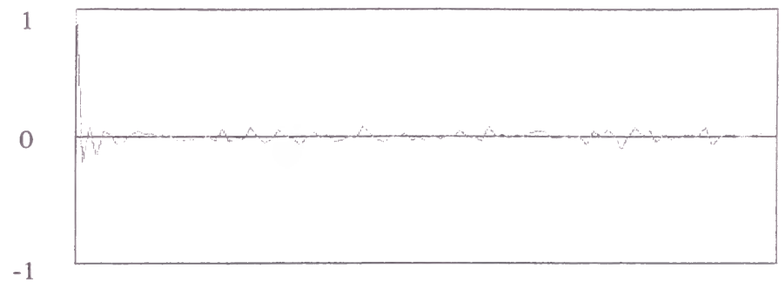


図 4.18: 自己相関関数

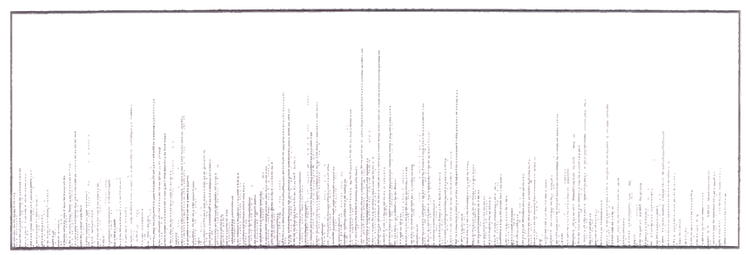


図 4.19: パワースペクトル

して，(1) 自己相関関数，(2) パワースペクトル，(3) リャプノフ指数，の三つを求めたところ以下のようになり，この系はカオス [54] の性質を有していると総合的に判断される．

(1) 自己相関関数

この時系列の自己相関関数を図 4.18に示す．明らかに急速に 0 に収束していることがわかる．

(2) パワースペクトル

この時系列のパワースペクトルを図 4.19に示す．図において，縦軸の周波数成分において明確なピークはないことがわかる．

(3) リャプノフ指数

最大リャプノフ指数 λ_1 を計算したところ，$\lambda_1 \approx 0.342$ となり正の値となった．

また，この系の性質を知るために，この系の制御パラメータに相当するファジィラベルの設定値を動かしたとき，系の振る舞いがどう変化するかを観測するための分岐図を図 4.20

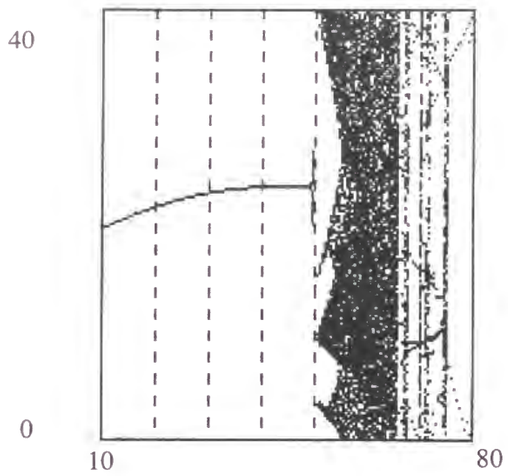


図 4.20: 単一ループから構成されるファジィ記号力学系の分岐図

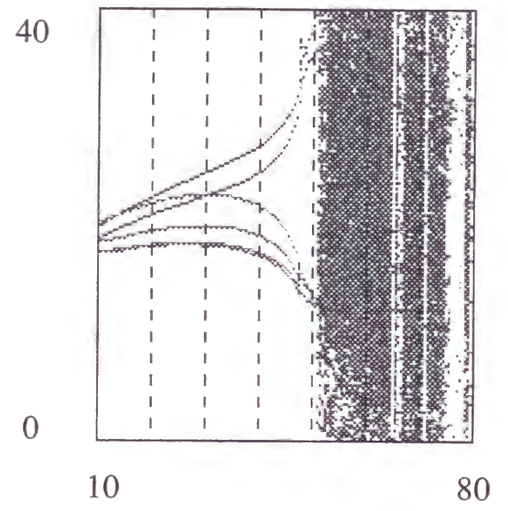


図 4.21: 複数のループから構成されるファジィ記号力学系の分岐図

（楕円制約が 1 つ，ループが 1 つの場合）と図 4.21（楕円制約が複数，ループが 2 つの場合）に示す．この分岐図において，(1) 周期解領域，(2) 遷移領域，(3) カオス領域など，パラメータ設定によって極めて多様な振る舞いの出現することが確認された．

[ラベル記号列の解析]

このようなカオス現象の要因を知るために，ファジィ記号力学系がカオス的挙動を示すときに選択されたファジィラベルの記号列を定性的に調べる．具体的には，図 4.14における

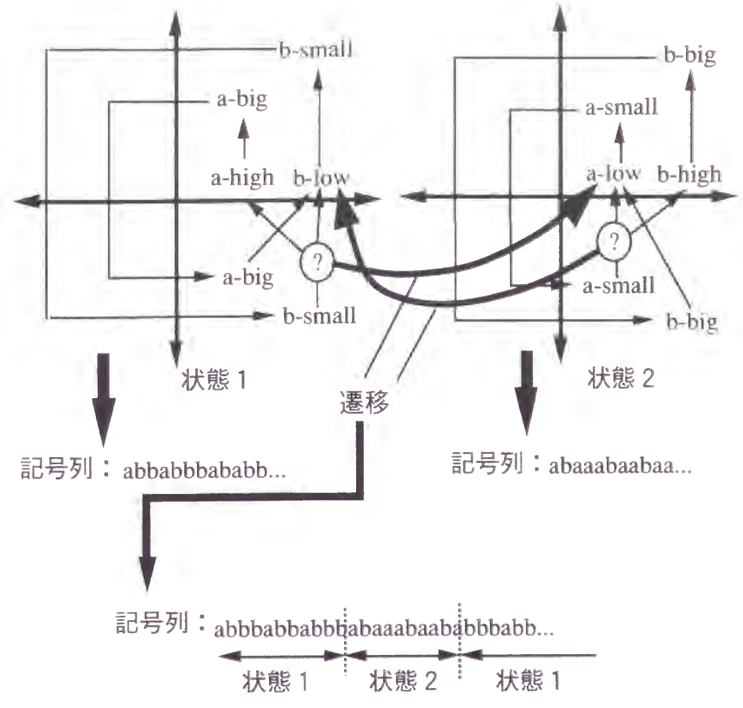


図 4.22: ラベル記号列の解析に基づく状態遷移図

ファジィラベル ‘small’ や ‘large’ を用いて ‘a-small’ や ‘b-big’ などの定性的な状態を定義することにより，図 4.22に示すような状態遷移図を得ることができる．この図においては，不安定アトラクタに対応するループが 2 つあり，しばらく 1 つのループ内を状態遷移した後，入力区間のわずかな違いによって別のループに切り替わるということが定性的に捉えられる．すなわち，この系でのカオス現象はファジィラベル選択の際のゆらぎによる状態遷移の積み重ねによって生み出されると考えられる．

このような構造は，カオス現象の代表的なモデルの一つであるローレンツの気象モデルの振る舞いと類似しており，極めて興味深い．つまり，どちらの時系列でも，2 つの不安定な不動点が存在し，一方の不動点の近くから徐々に発散し，別の不動点付近に切り替わる様子が見られる．

このようなファジィ記号力学系においてカオス的な挙動が生み出される要因としては，下位層である制約伝播力学系の安定的な振舞いに，上位層からファジィラベルの選択という働きかけ（ゆらぎ）が付加されたことが考えられる．

4.3.3 考察

プロセス指向の解概念

このような制約伝播プロセスによるアプローチでは，注目する制約および変量を次々に変えることにより，絶えず様々なパターンを次々に作り出しており，全体的整合性は満たされていないかもしれないが，極めて多様な解を常に探索していると考えられる．そこで導出される解は，解自体が動的な意味をもっており，従来の収束解を包含した形での新しいプロセス指向の解概念と捉えることもできる．すなわち，従来の最終解である 1 周期解だけでなく振動するような周期解も動的なあるいは確率的な解として考えること意味している．また見方を変えれば，これは問題を直接緩和するのではなく，解概念の方を従来より緩和していることになり，ある意味で現実の社会でもしばしば見受けられることと考えられる．

ファジィ記号力学系の活用

前章で提案した階層型問題解決システムとの関連させて考えると，上で述べた制約伝播力学系およびファジィ記号力学系は下位層での処理に対応し，上位層では全体的整合性の充足を図る記号処理が行われると捉えることができる．すなわち，下位層でのファジィ記号力学系がカオス的挙動を示すことで多様な解を探索し，その振る舞いを記号列という形で上位層に伝えるとともに，上位層はそれらの記号列を用いて全体として整合する解を探索することが考えられる．その際，上位層から下位層への働きかけとして，下位層での時間的な探索プロセスを制御する系のパラメータ（記憶係数など）の更新なども考えられる．

GA による制約充足問題の解法との融合

上で述べた上位層での処理としては，前章で採用した併合法 (invasion) などの厳密解法だけでなく，より大規模な問題に有効な近似解法の導入も考えられる．とくに進化プロセスを模倣した遺伝的アルゴリズム (GA) を上位層での制約充足問題の解決に用いることを考えると，下位層での制約伝播プロセスによる解の探索と併用することが可能となる．すなわち，下位層での制約伝播プロセスでは，局所的には整合している部分解を次々と求めることができるため，それらを上位層での GA プロセスに対して適当なタイミングで導入することが考えられる．もちろん，初期解として導入することもでき，多様な部分解を GA プロセスに与えられるので多様性維持の効果も期待できる．

4.4 結言

本章では，前章で提案したように，ネットワーク内の全てのリンク構造を同時並行的に活用し，自己組織的・自律分散的に問題解決を行うのではなく，一度に一つのリンク構造だけを選びそれらを逐次的にたどる制約伝播による問題解決プロセスについて検討を行なった．このとき，伝播されるのは区間（制約区間）であり，制約伝播の際に連続量の区間を記号化せずにそのまま伝播する制約伝播力学系においては，安定平衡点に収束する極めて安定的な振る舞いが得られることを数学的に示すとともに，計算機シミュレーションにより確認を行なった．また，制約伝播の際に連続量を記号化（符号化）する手段としてファジィネスを導入したファジィ記号力学系においては，カオス的現象を生み出すような複雑性が内包されていることを計算機シミュレーションを通して明らかにした．

このような複雑な現象が生み出される要因としては，制約伝播力学系の安定的な振舞いに，ファジィラベルの選択という働きかけ（ゆらぎ）が付加されたことが考えられる．つまり，制約伝播の際に記号化のためにファジィラベルの選択のプロセスが介入しているからである．このことは，記号力学の立場から，選択されるファジィラベルの記号列を調べることによっても明らかにした．一般に，記号化を通して力学系などの連続量の状態遷移を構造的に把握する分野は記号力学と呼ばれるが，本章ではファジィネスの介入した記号力学系の複雑性の一端を示したといえる．

さらに，前節で考察したように，このような制約伝播プロセスによるアプローチでは，極めて多様な解（全体的整合性は満たされていないかもしれないが）を常に探索しており，そこで導出される解は，解自体が動的な意味をもっており，従来の静的な解を包含した形での新しいプロセス指向の解概念と捉えることもできる．また，このようなアプローチを前章で提案した階層型問題解決システムに導入することについても検討し，遺伝的アルゴリズムを用いた上位層での制約処理との融合・併用の可能性について考察を行なった．

第 5 章

連続値入出力を扱う強化学習

5.1 緒言

近年，自律エージェントの学習法として，環境との相互作用を通して環境に適した行動を試行錯誤的に獲得する強化学習が注目されている．強化学習における目的は，将来にわたり獲得する報酬の和を最大化するように，状態から行動への写像（政策）を学習することである．強化学習は古くから心理学の分野で研究されてきたが，工学の分野では初期の人工知能研究における Samuel のチェッカープログラムにまで遡ることができる．強化学習のアルゴリズムとしては，確率的学習オートマトン [57]，$TD(\lambda)$ 法 (Temporal Difference 法) [58]，Q-Learning [59] や，遺伝的アルゴリズムの分野で考案されバケツリレー法 [61] および Profit Sharing 法 [62] などをはじめとして様々な手法がこれまでに提案されている．

これらの強化学習のアルゴリズムの中でよく用いられるものとして，Watkins により提案された Q-Learning [59] がある．しかし，この手法は離散的な状態・行動を対象としており，現実の複雑な環境に対応するためには，連続値を含めた多様な入出力データを扱う必要がある．本研究では，この Q-Learning において連続値をもつ状態・行動を取り扱うことを可能にするために，ファジィ推論を導入したファジィ内挿型 Q-Learning [76] を提案する．さらにこの学習法に対して，前件部を含むファジィルールの更新（学習）や離散的な行動に対する学習の効率化などによる改良を加え，二種類の制御問題への適用を通して他の従来手法と比較することによってその有効性を検証する．

以下，**5.2 節**では強化学習の概要について説明した後，**5.3 節**において我々が提案するファジィ内挿型 Q-Learning についてアルゴリズムや特徴を詳しく述べる．また，行動が離散的である問題に対する学習の効率化のための手法についても言及する．さらに **5.4 節**では，これらの提案手法を倒立振子制御問題と大型船操舵問題の二種類の制御問題へ適用し，他の従来手法と比較することによってその有効性を明らかにする．

5.2 強化学習の概要

5.2.1 強化学習の枠組みと特徴

自然界における生物は，未知環境において報酬 (食物など) を獲得し罰 (天敵など) から逃れるような適切な行動を，試行錯誤によって獲得するという基本的な適応能力を有している．強化学習はこのようなシステムを工学的に模倣した枠組みである [79].

強化学習の基本的な枠組みは，第 2 章の図 2.6に示した．学習者は環境からの感覚入力や学習者の内部状態, あるいはそれらの組合せによって表わされる状態集合 S の中から現在の状態 s_t を認識し，行動集合 A の中から行動 a_t を一つ決定し，次の状態 s_{t+1} に遷移する．行動は状態の関数 $a_t = f(s_t)$ で表され，関数 f は政策 (policy) と呼ばれる．このとき状態 s_t での行動 a_t に対して強化信号 (報酬)r_t が与えられる．この強化信号は，与えられた直前の行為の決定のみに対してではなく，その状態への遷移に関与した行為の系列についての評価に基づいて与えられる．

学習の目的は，次式で表される現在から未来にわたる報酬の減衰総和 V_t を最大化することにある．

$$V_t = \sum_{n=0}^{\infty} \gamma^n r_{t+n} \tag{5.1}$$

ここで，γ は割引率を表し，将来の報酬をどの程度考慮するかを決定するパラメータである．もし $\gamma = 0$ であれば，現在得られる報酬のみに着目し，未来にどのような報酬が得られるかについては全く考慮に入れないことになる．逆に $\gamma = 1$ のときは，得られた報酬の時間減衰はなく，即座に得られる報酬も将来に得られる報酬も大きさが同じであれば全く同じ価値を持つと見なすことになる．このとき学習者の獲得する政策は, 先ほどとは反対に, 遠い将来に得られる報酬も考慮に入れたものとなる．つまり，γ の値の大小によって，どのくらい先の未来までを考慮するかが決まる．

しかし，未来に得られる報酬は，行為を決定する時点においては推測することはできない．よって，一般には過去から現在までの報酬の重み和

$$\hat{V}_t = \sum_{n=0}^{t} \gamma^{t-n} r_n \tag{5.2}$$

を V_t の近似値として利用し, これについての最大化を学習の目的に置き換える．

強化学習の特徴としては，環境や学習者自身に関する先験的知識をほとんど必要とせず，学習者の経験によって与えられたタスクを遂行するような目的行動を学習する教師なし学習であることが挙げられる．また，環境からの報酬は即座には与えらず，時間遅れを伴って与

えられる．すなわち，報酬は個々の行動の善し悪しを教示するわけではなく，一連の行動の結果に対して評価を与えるものである．そのため強化学習は最終的な結果の良し悪しは既知であるが，途中の各行動に対しての良し悪しはその時点では判断できないような問題に対して有効である．

5.2.2 強化学習の主な実現手法

$TD(\lambda)$ 法

$TD(\lambda)$ 法は R.S.Sutton によって考案された一種の時系列予測手法である．TD の目的は，状態 x において最適な行動をとることによって獲得される将来の報酬の総和を学習を通して予測することである．状態 x における将来に得られる報酬の重み和

$$V_t(x) = \sum_{i=t}^{\infty} \gamma^{i-t} \cdot r_i \tag{5.3}$$

を効用 (utility) と呼び，効用の予測値 $\hat{v}(x)$ をつぎのように見積もることによって学習を行う．

時刻 t において状態が x_t であって行動を実行した結果，状態は x_{t+1} に遷移し，強化信号 r が得られたとき，効用の予測値 $\hat{v}(x)$ の期待値は

$$\hat{V}'(x_t) = r_t + \gamma \hat{V}(x_{t+1}) \tag{5.4}$$

である．したがって，予測誤差

$$e = \hat{V}'(x_t) - \hat{V}(x_t) \tag{5.5}$$

を 0 に近づけるように $\hat{v}(x)$ を更新させる．この予測誤差は “TD-error” と呼ばれる．状態 x_t における効用の予測誤差 e は，x_t に到達可能なすべての状態 x の効用の予測値に影響を及ぼすことから，過去に通ってきたすべての状態 $x_{t'}(t' \leq t)$ における効用を次式で更新する．

$$\hat{V}(x_{t'}) \leftarrow \hat{V}(x_{t'}) + \alpha e \lambda^{(t-t')} \tag{5.6}$$

ここで α は学習率，λ は時刻による予測誤差の減衰率であり，$0 \leq \alpha, \gamma \leq 1$ の定数である．減衰率 λ は過去の状態系列に対する予測誤差の伝播率を表しており，値が大きいほど現在得られた経験を過去の状態系列にさかのぼって反映させる．

$\hat{V}(x)$ は状態に対する評価値であり，行動決定は政策 (policy) に基づいて実行される．政策は，状態 x において行動 a を実行した場合の効用の予測値 $Policy(x, a)$ が用いられる．政策の学習は，状態の効用の学習と同時に同じ予測誤差を利用して行われる．状態 x_t におい

て行動 a_t を選択した場合に，現在の状態に至った過去の状態と行動の系列 $(x_{t'}, a_{t'})(t' \leq t)$ に対して

$$Policy(x_{t'}, a_{t'}) \leftarrow Policy(x_{t'}, a_{t'}) + \alpha e \lambda^{(t-t')} \tag{5.7}$$

により更新する．

Q-Learning

Q-Learning は，Watkins によって提案された強化学習の実現方法の一つである．$TD(\lambda)$ では評価値として状態に対する評価を見積もるのに対し，Q-Learning では状態と行動の組に対する評価を見積るものである．この評価を Q 値と呼び，状態と行動の組から評価の見積りを導く関数を Q 関数と呼ぶ．Q 値は，$TD(0)$ における $Policy(x, a)$ に相当し，効用 $V(x)$ を介さずに直接見積もられる．そのため，このような Q-Learning は 1-Step Q-Learning とも呼ばれる [60]．強化学習システムは学習要素と行為決定要素に分けることができる．以下に，Q-Learning における各要素の実現方法を述べる．

(1) 学習要素

Q-Learning では状態と行動の組に対して Q 値と呼ばれる評価の見積りを行って学習を進めるものであり，その Q 値をもとに行動決定を行う．したがって，強化学習システムにおける学習要素は，Q-Learning では各状態と行動に対して Q 値の見積りを行う要素に他ならない．Q 値の見積りは以下のように行われる．

時刻 t において状態が x_t であって行動 a_t を実行した結果，状態は x_{t+1} に遷移し，強化信号が得られたとき，$Q(x_t, a_t)$ の期待値 Q' は

$$Q'(x_t, a_t) = r_t + \gamma V(x_{t+1}) \tag{5.8}$$

$$V(x_{t+1}) = \max_b Q(x_{t+1}, b) \tag{5.9}$$

である．Q-Learning では，この期待値 Q' に Q 値を近づけるように更新することによって学習を進行する．この操作を行う Q 値の更新式を以下に示す．

$$Q(x_t, a_t) \leftarrow (1-\alpha)Q(x_t, a_t) + \alpha(r_t + \gamma \max_b Q(x_{t+1}, b)) \tag{5.10}$$

α は $0 < \alpha \leq 1$ なる定数であり，学習率と呼ばれる．上式は，次のステップで最適と思われる行動を選択したときに得られると見込まれる評価の見積もり $\max_b Q(x_{t+1}, b)$ を割引率 γ だけ割り引いた値と，そこで直接得られた強化信号 r_t の和で表される期待評価にもとの Q 値 $Q(x_t, a_t)$ を近づける．このとき，学習率はこの値に近づける度合いを表しており，大きいほど急激に近づけることになる．したがって学習率が大きいほど学習がはやく進むと思われるが，状態変化の不確実さが存在するなどシステムの状態や行動が複雑な場合は学習率が大きいほど Q 値が収束しにくくなる可能性がある．よって，感覚入力や報酬にノイズが含まれるなど学習環境が複雑な場合は学習効率やノイズの影響の最小化を考慮して学習率を適切に設定する必要がある．Q-Learning では，式 (5.10) で Q 値を更新し続けることにより，$\max_b Q(x_{t+1}, b)$ を最適な行動を取り続けたときの強化信号の重み和 v_t に近づけることができる．

(2) 行動決定要素

Q-Learning では，学習された Q 値をもとに行動を決定する．行動決定の方法として Watkins は次式のような Boltzmann 分布に基づく確率的な行動選択法を提案した．

$$P(a|x) = \frac{\exp(Q(x,a)/T)}{\sum_{b \in \text{possible actions}} \exp(Q(x,b)/T)} \tag{5.11}$$

上式を状態 x_t において行動 a を選択する確率として定義し，この確率をもとに学習者は行動を決定し探索を行う．ここで上式の T は温度定数であり，この値が大きいほど各行動に対する Q 値の差が確率に反映されないので行動はランダムになり積極的に探索を行うことになる．一方，T が小さいほど探索が行われにくくなり，極限では Q 値を最大にするような行動が選ばれる戦略 (greedy policy) をとる．したがって，学習の初期では T を大きくして最適な行動系列の探索を行い，学習が進むにつれて小さくしていく戦略が有効である．

Q-Learning は，状態遷移確率が直前の状態にだけ依存し，それ以前の経過に無関係である性質をもつマルコフ決定問題に対して，以下の収束に関する定理が成り立つ．

定理 (Q-Learning の収束性) 有限範囲内の強化信号 $|r_t| \leq R$ が与えられる状況において，すべての実行可能な状態と行動の組 (x, a) に対して無限回 Q 値が計算され，学習率が $0 \leq \alpha_t < 1$ かつ

$$\sum_{t=1}^{\infty} \alpha_t(x,a) = \infty, \quad \sum_{t=1}^{\infty} [\alpha_t(x,a)]^2 < \infty, \quad \forall x, \forall a \tag{5.12}$$

の条件で t の増加に従って適切に減少するなら，Q-Learning により生成される列 $\{Q_t(x,a)\}$ は，全ての組 (x,a) に対して確率 1 で $Q^*(x,a)$ に収束する．

Q-Learning では，学習要素および行動決定要素において Q 値という 1 つの評価基準に基づいて学習し，新たに積むべき行動を決定している．このため，学習率のパラメータである

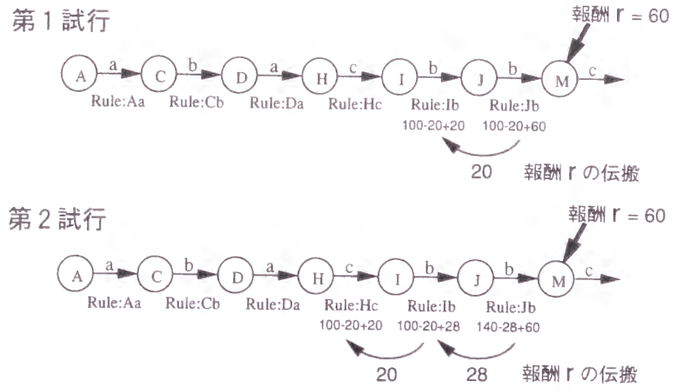


図 5.1: バケツリレーアルゴリズム

学習率および割引率，行動決定要素のパラメータである温度定数は，学習のための経験にいづれも大きな影響を及ぼす．したがって，学習をより効率的にするためには，状態空間の性質と，強化信号の性質および問題の成功条件，学習に必要な探索の深さを考慮してパラメータを設定する必要がある．

分類子システム（バケツリレー法，Profit Sharing 法）

強化学習の枠組みが組み込まれたものとして，Holland が提案した遺伝的アルゴリズムに基づく分類子システム (Classifier System) がある．分類子システムでは，実行要素として if–then 形式のプロダクションルールが用意される．各ルールには「強さ」が割り当てられており，条件部が満足された実行可能ルール間の競合は強さに基づいて解消される．学習者がある程度行動を繰り返し，その経験に基づいてルールの強さを変更していき，十分な変更がなされた時点で，強さを適合度と見なして遺伝的アルゴリズム (Genetic Algorithm) を適応する．これにより，無駄なルールは淘汰され，有効なルールを組み合わせて新たなルールが生成される．

ルールの強さの更新時における報酬割当て (Credit Assignment) の方法には，バケツリレー法 (Bucket Brigade Algorithm) と Profit Sharing 法（利益共有法）がある．

(1) バケツリレー法

バケツリレー法では，行動を選択する毎に一種の賭を行う．あるステップで活性化したルールは，一定の賭金を支払う．その競合ルールの中からルーレットにより選択されたルー

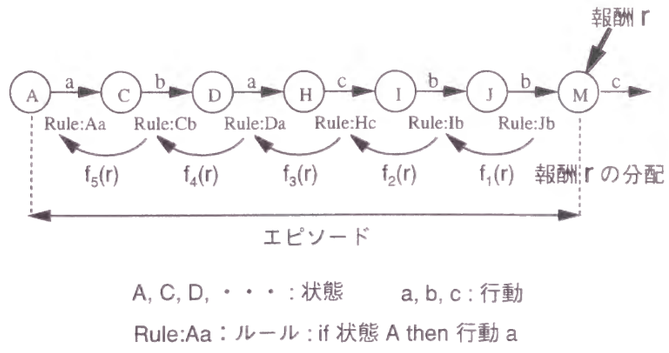


図 5.2: Profit Sharing 法

ルは勝者と見なされ，そのルール実行結果として得られた報酬と次のステップの賭金の合計が与えられる．すなわち，ある時間ステップ t においてルール R_i が発火し，次の時間ステップ $t+1$ においてルール R_j が発火したとすると，ルール R_i の強さ S_i は以下のようにして更新される．

$$S_i(t) \leftarrow (1-b)\, S_i(t) + b\, S_j(t) \tag{5.13}$$

ここで，b は $0 < b < 1$ であり学習率に相当する．またある時間ステップでルール R_i が発火し，その結果報酬 r が得られた場合は次式にしたがって更新される．

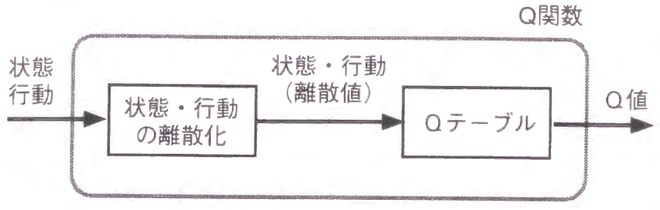
$$S_i(t) \leftarrow (1-b)\, S_i(t) + r \tag{5.14}$$

バケツリレー法では，得られた報酬は，ただちに過去のルールに伝搬されずに，次の実行の際に一段階だけ伝搬する．この点は，Q-Learning の場合の Q 値の変更における報酬の伝搬と同様であり，学習の速度は遅い．

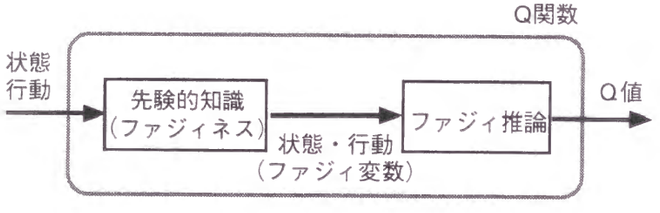
(2) Profit Sharing 法

Profit Sharing 法（利益共有法）では，実行されたルールの履歴を保存しておき，報酬が得られるたびに，報酬の値を減衰させながら過去にさかのぼってルールの強さを更新する．報酬が得られてから次の報酬が得られるまでのルールの選択系列はエピソード (episode) と呼ばれ，エピソードに参加したルールに得られた報酬を分配する．すなわち，ある時間ステップ t において報酬 r_t が得られるとき，エピソードに参加したルール R_{t-i} の強さは次式のように更新される．

$$S_i(t) \leftarrow (1-b)\, S_i(t) + b\, f_i(r) \tag{5.15}$$



(a) Q-Tableを用いたQ関数（従来手法）



(b) ファジィ推論法を用いたQ関数（提案手法）

図 5.3: 従来手法と提案手法の枠組み

ここで，b は学習率 $(0 < b < 1)$，r はエピソードの最後において得られた報酬であり，$f_i(r)$ はエピソードの最後のステップ t から数えて i ステップ前のルールに分配する報酬の大きさを決定する関数で強化関数と呼ばれる．(Fig.1 参照)．

Profit Sharing 法では，過去に実行されたルールの強さも一括して変更するため，学習の速度は速い．この Profit Sharing 法は，学習の効率性を重視し報酬を獲得した経験を積極的に強化するため，経験強化型の代表的な手法の 1 つである．

5.3 ファジィ内挿型 Q-Learning の提案

5.3.1 提案手法の枠組み

上記のような従来の Q-Learning は主に離散的な状態・行動を対象としており，現実の複雑な環境に対応するためには，連続値を含めた多様な入出力データを扱う必要がある．そのための手法としてはこれまでに，小脳のモデルである CMAC を用いて Q 関数を表現する方法 [63] や後述の階層型ニューラルネットワークを用いて Q 関数を表現する Q-net [64] などが提案されている．それに対して我々は，先験的知識を導入することが容易で学習速度の高速化も期待できる手法として，ファジィ推論を用いて Q 関数を表現することにより，連続値の状態・行動を取り扱うことを可能にするファジィ内挿型 Q-Learning を提案する．

従来の Q-Learning ではすべての状態と行動の組合せについての表 (Q-Table) を参照する

表 5.1: ファジィ内挿型 Q-Learning のアルゴリズム

Step1. 現在の状態 x_t を観測する

Step2. ファジィ推論により Q 値 $Q(x_t, a)$ を求める

Step3. Q 値に基づいて行動を選択する

Step4. 選択された行動 a_t を実行し，次状態 x_{t+1} を観測する

Step5. ファジィ推論により Q 値 $Q(x_{t+1}, b)$ を求める

Step6. Q 値の更新幅 $\Delta Q(x_t, a_t)$ を計算する

Step7. Q 関数（ファジィルール）の更新を行う

Step8. **Step1** にもどる

ことにより，Q 値を出力していたのに対して，提案手法では状態・行動をファジィ変数としてファジィ推論により Q 値が出力される（図 5.3参照）．すなわち，Q-Learning において見積もられる Q 関数をファジィ推論を用いて近似することによって，連続値入出力を含む問題における学習を行う．

5.3.2 提案手法のアルゴリズム

まず表 5.1に，ファジィ内挿型 Q-Learning のアルゴリズムの概略を示す．アルゴリズムの大まかな流れは従来の Q-Learning とほぼ同じであるが，Q 関数の近似にファジィ推論を導入するために，**Step2** および **Step5** におけるファジィ推論を用いた Q 値の導出，**Step3** における連続値行動を出力するための行動選択，ならびに **Step7** におけるファジィルールの更新について新たな手法が加わっている．それらの各ステップについて以下に詳しく説明する．

ファジィ推論による Q 値の導出

本手法では，状態と行動の組に対し，関数型推論法 [65] を用いたファジィ推論によりその Q 値を導出する．いま学習者が環境の状態として n 次元連続値ベクトル $\boldsymbol{x} = (x_1, \ldots, x_n)$ を観測し，行動として連続値 a を出力するとする．このとき，状態 x，行動 a から Q 値を導出するためのファジィルールは以下のようになる．

$$R_i : \text{If } x_1 \text{ is } B_{i1} \text{ , } \cdots \text{ , } x_n \text{ is } B_{in} \text{ , } a \text{ is } B_{ia} \text{ then } Q = f_i(x_1, \cdots, x_n, a), \tag{5.16}$$

ここで，$B_{ix}, \ldots, B_{ia}$ はファジィ集合である．Q 値は，以下のように，各ファジィルール R_i について適合度 ω_i を計算し，これらの重み付き平均として推論結果を非ファジィ化することによって導出される．

$$\omega_i = B_{ix_1}(x_1) \cdot \cdots \cdot B_{ix_n}(x_n) \cdot B_{ia}(a) \tag{5.17}$$

$$Q(x_1, \cdots, x_n, a) = \frac{\sum \omega_i \cdot f_i(x_1, \cdots, x_n, a)}{\sum \omega_i} \tag{5.18}$$

また，ファジィルール R_i の前件部のメンバーシップ関数および後件部の関数としては一般に任意の関数を用いることができるが，以下のガウス基底関数 $g_{ij}(x_j)$ および線形式 $f_i(\boldsymbol{x}, a)$ を用いるものとする．

$$g_{ij}(x_j) = \exp\{-\frac{1}{2}\Big(\frac{x_j - \mu_{ij}}{\sigma_{ij}}\Big)^2\} \tag{5.19}$$

$$f_i(\boldsymbol{x}, a) = C_{i1} \cdot x_1 + \ldots + C_{in} \cdot x_n + C_{ia} \cdot a + C_i \tag{5.20}$$

ここで，(μ_{ij}, σ_{ij}) はファジィルール前件部のパラメータ，(C_{ij}, C_i) はファジィルール後件部のパラメータである．

行動選択

ファジィ内挿型 Q-Learning は連続値の出力（行動）を扱うものであるため，従来の手法のようにとり得るすべての行動に対する Q 値を求めるのは困難である．したがって，まず状態 x_t において有限個の行動 a_i に関して Q 値を求めた後，これらの Q 値に対し Boltzmann 分布（式 (2) 参照）に基づくスケーリングを行う．基本的にはこの Q 値の分布に従った確率で行動を選択する必要がある．

このための手法としては，モンテカルロ法を用いることによりその分布に従った確率で行動を選択することが考えられる．しかし，この方法では無駄な試行が多数発生するので行動決定のための計算効率が悪い．そこで，このモンテカルロ法に基づく手法を簡略化したものとして，Q 値の分布に基づいて連続値の行動を選択できるように拡張したルーレット選択を考える（図 5.4参照）．この方法ではまず，有限個の行動 a_i に関する（スケーリングされた）Q 値に基づいてルーレットを作る．つぎに，ルーレット選択が乱数によって行われ区分 j（a_i と a_{i+1} の間）が選択されるとき，区分 j 内において乱数によって決定された位置に基づいて区間 j 内の連続値の行動を選択する．

ただし，このような拡張ルーレット選択法の他にも，ある一定の確率は貪欲法（Greedy 法）で行動選択し，残る確率でランダムに行動する手法や，eligibility を用いる Exploration-Exploitation Policy [69] などの方法により行動を選択することも考えられる．

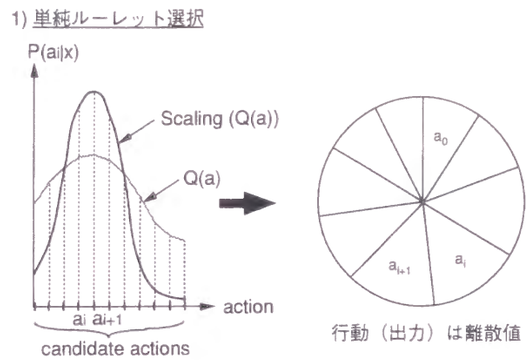


図 5.4: 行動選択方法

Q 関数の更新

提案手法では，**Step6** において $Q(x_t, a_t)$ の更新幅 $\Delta Q(x_t, a_t)$ を求めた後，それをもとに Q 関数（ファジィルール）の更新を行う．すなわち，ファジィ推論により導かれる値が，期待評価 $Q'(x_t, a_t)(= Q(x_t, a_t) + \Delta Q(x_t, a_t))$ に近づくように，Q 値を導出するのに用いられたファジィルールのパラメータ $(\mu_{ij}, \sigma_{ij}), (C_{ij}, C_i)$ を変更する．その際，次式で定義される二乗誤差を最小化する方向でパラメータ値を変更し，そのための手法として最急降下法を用いる．

$$\begin{aligned} E &= \frac{1}{2}(\Delta Q)^2 \\ &= \frac{1}{2}\{r_t + \gamma \max_b Q(x_{t+1}, b) - Q(x_t, a_t)\}^2 \end{aligned} \tag{5.21}$$

この評価関数の勾配（パラメータに関して偏微分）を求めることにより，ファジィルール前件部および後件部それぞれについて評価関数 E を減少させる以下のような学習則が導かれ

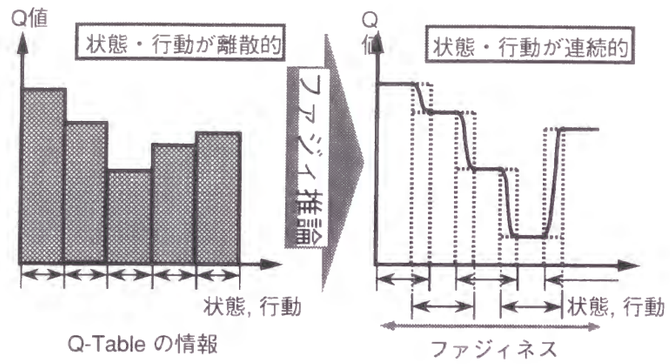


図 5.5: 提案手法の特徴

る．すなわち，以下の学習則により，ファジィルールの各パラメータ値が更新される．

$$\begin{cases} \mu_{ij} \leftarrow \mu_{ij} + \varepsilon_a \frac{1}{\sigma_{ij}} \frac{x_j - \mu_{ij}}{\sigma_{ij}} W_i (f_i - Q) \Delta Q \\ \sigma_{ij} \leftarrow \sigma_{ij} + \varepsilon_a \frac{1}{\sigma_{ij}} \left(\frac{x_j - \mu_{ij}}{\sigma_{ij}} \right)^2 W_i (f_i - Q) \Delta Q \end{cases} \tag{5.22}$$

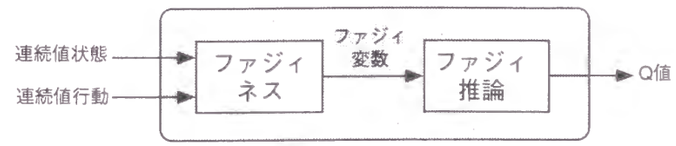
$$\begin{cases} C_{ij} \leftarrow C_{ij} + \varepsilon_c W_i x_j \Delta Q \\ C_i \leftarrow C_i + \varepsilon_c W_i \Delta Q \end{cases} \tag{5.23}$$

ここで，$W_i = \omega_i / \sum_i \omega_i$ であり，$\varepsilon_a, \varepsilon_c$ はそれぞれファジィルール前件部および後件部に対する学習率である．

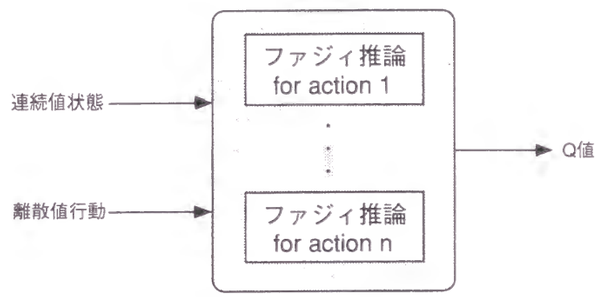
5.3.3 提案手法の特徴

提案手法は，状態 x と行動 a から Q 値を導出する際にファジィ推論を用いるものであり，ファジィルールの表現形式により Q 関数を近似している．したがって，従来の Q-Table を用いる方法では状態・行動をともに離散的に扱うため，Q 関数の曲面は不連続な柱状分布を成していたものが，ファジィ推論の導入によりなめらかに内挿（補間）することが可能になり，汎化能力が期待できる（図 5.5参照）．

連続状態空間を離散化せずに Q-Learning を用いる方法としては，Lin らが提案した Q-net [64] がある．これは階層型ニューラルネットワークを用いて Q 値を算出するものであり，Q 関数の表現法がファジィルールによる表現を用いている我々の提案手法とは異なるが，最急降下法が用いられている点は共通している．しかしながら，Q-net では取り得る各行動に対してそれぞれネットワークを用意する必要があり，このような構成は $OAON$ アーキテクチャ(One Action One Network) と呼ばれる．このような Q-net では，状態は連続値として扱えるが行動は離散的にしか扱えないのに対して，本章で提案する方法では行動も連続値をとることが可能である．



(a) 連続値行動の場合



(b) 離散値行動の場合

図 5.6: 離散値行動に対する $OAON$ アーキテクチャ

また，先験的知識としてファジィルールのメンバーシップ関数の構造と初期設定が与えられているため，同じくパラメータ学習則として最急降下法がを採用している Q-net に比較して学習速度が速いことが考えられる．なお，このメンバーシップ関数の前件部についても後件部と同様に最急降下法により学習がなされる．さらに，本章ではメンバーシップ関数の後件部パラメータ (C_{ij}, C_i) は全て初期値ゼロにして例題への適用を行うが，もし対象としている問題領域に関して人間の先験的知識（領域知識）が部分的にでも利用できる場合は，それを用いて初期値を設定することが可能でありより一層の学習の高速化が期待できる．

5.3.4 離散値行動に対する学習の効率化

前節で述べたように，Lin らが提案した Q-net では，行動は離散値のみを扱うものとし，有限個の取り得る各行動に対してそれぞれニューラルネットワークを用意することにより学習を行っていた．このような $OAON$ アーキテクチャは学習の効率化のために有効であると考えられる．

ファジィ内挿型 Q-Learning においても，行動が離散値である場合にはこの $OAON$ アーキテクチャのように，それぞれの行動ごとにファジィ推論システムを用意することが考えら

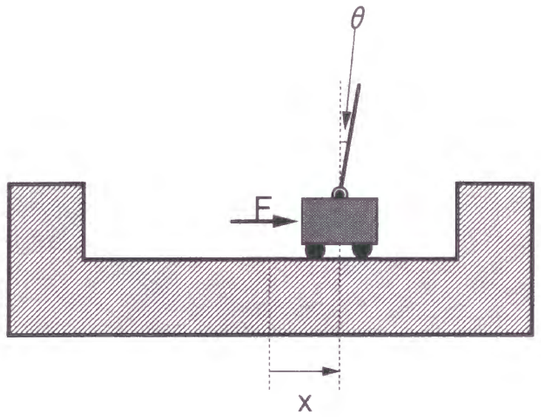


図 5.7: 倒立振子制御問題

れる（Fig.4 参照）．すなわち，各離散値行動に対して，以下のように前件部に状態 x のみを含み行動 a は含まないファジィルールを用意し，それらを用いてファジィ推論を行い Q 値を導出する．

$$R_i \ : \ \text{If } x_1 \text{ is } B_{i1} \ , \ \cdots \ , x_n \text{ is } B_{in} \text{ then } Q = f_i(x_1, \cdots, x_n, a), \tag{5.24}$$

また，行動選択については行動が離散値なので従来の Q-Learning と同様に，Boltzmann 分布（式 (2) 参照）に基づく確率的な行動選択を行う．

このような簡略化アーキテクチャを採用することにより，Q 関数を同定する問題が部分問題に分割されるため学習の効率化が期待できる．ただし，行動が連続値であり離散値として扱えない場合にはこの手法は用いることができない．

5.4 制御問題への適用

本章では，前章で提案したファジィ内挿型 Q-Learning を倒立振子制御問題と大型船操舵問題の二種類の制御問題へ適用し，他の従来手法と比較することによってその有効性を検証する．

5.4.1 倒立振子制御問題への適用

倒立振子制御問題では，振子の角度 θ, 角速度 $\dot{\theta}$, 台車の位置 x, 速度 $\dot{x}$ を観測量とし，行動として台車に加える水平外力（操作量）F を決定する（図 5.7参照）．倒立振子系を表す微

表 5.2: 倒立振子制御問題における状態分割

状態変数	区間	分割数
$\theta[^\circ]$	$[-12,-6]$ $[-6,-1]$ $[-1,+1]$ $[+1,+6]$ $[+6,+12]$	5
$\dot{\theta}[^\circ/\mathrm{s}]$	$[-\infty,-50]$ $[-50,+50]$ $[+50,+\infty]$	3
$x[\mathrm{m}]$	$[-2.4,-1.2]$ $[-1.2,-0.4]$ $[-0.4,+0.4]$ $[+0.4,+1.2]$ $[+1.2,+2.4]$	5
$\dot{x}[\mathrm{m/s}]$	$[-\infty,-0.5]$ $[-0.5,+0.5]$ $[+0.5,+\infty]$	3

分方程式はつぎのように記述することができる [70].

$$\ddot{\theta}_t = \frac{g\sin\theta_t + \cos\theta_t\left[\frac{-F_t - ml\dot{\theta}_t^2\sin\theta_t + \mu_c\mathrm{sgn}(\dot{x}_t)}{m_c+m}\right] - \frac{\mu_p\dot{\theta}_t}{ml}}{l\left[\frac{4}{3} - \frac{m\cos^2\theta}{m_c+m}\right]} \tag{5.25}$$

$$\ddot{x}_t = \frac{F_t + ml\left[\dot{\theta}_t^2\sin\theta_t - \ddot{\theta}_t\cos\theta_t\right] - \mu_c\mathrm{sgn}(\dot{x}_t)}{m_c+m} \tag{5.26}$$

本章では，時間ステップを 0.02 秒とし Runge-Kutta-Gill 法を用いて計算機シミュレーションを行う．また問題に関する各パラメータは重力加速度 g : $9.8[m/s^2]$, 台車の質量 m_c : $1.0[kg]$, 振子の質量 m : $0.1[kg]$, 振子の半分の長さ l : $0.5[m]$, 台車とレールの摩擦係数 μ_c : 0.0005, 振子とヒンジの摩擦係数 μ_p : 0.000002 とした．

学習エージェントには上記のダイナミクスは与えられておらず，振子が 12° 以上傾いたとき，あるいは台車がレールからはみ出したとき失敗したとみなし，強化信号（罰）-1 が与えられるものとする．

このような問題設定において，状態空間分割に基づく Q-Learning を適用した場合と Q-net を適用した場合，提案手法を適用した場合の比較を行う．それぞれの方法は次のような設定で適用した．なお，学習のパラメータとしては学習率 $\alpha = 0.5$，割引率 $\gamma = 0.99$，温度定数 $T = 0.005$ として学習を行った．

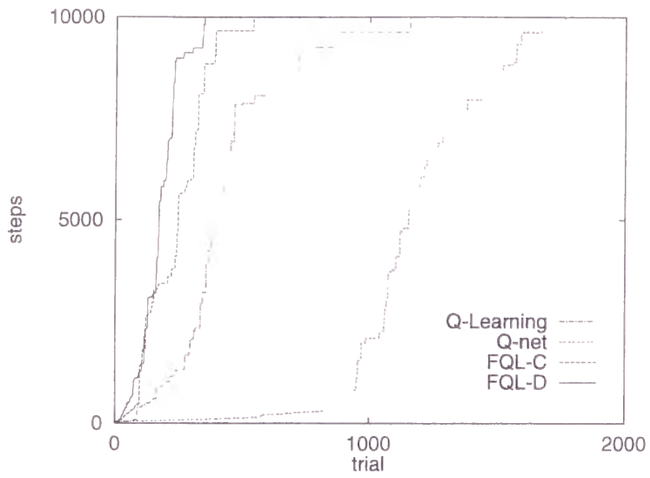


図 5.8: 倒立振子制御問題における学習速度の比較

(a) Q-Learning（離散値入力/離散値出力）：

表 5.2に示すように，状態空間を $5 \times 3 \times 5 \times 3 = 225$ 個に分割し，操作量として取り得る行動 F は $-10, 0, +10$[N] の 3 値とする．

(b) Q-net（連続値入力/離散値出力）：

行動は上述の 3 値の中から選択するものとし，それぞれの行動ごとに 16 ユニットの中間層をもつ 3 層の階層型ニューラルネットワークを用いて Q 値を出力する．なお，学習速度の向上のために慣性項を加えるとともに，1 回のステップでバックプロパゲーション (BP) をある一定回数行うなどした．

(c) ファジィ内挿型 Q-Learning 方法 1（連続値入力/連続値出力，FQL-C）：

行動を連続値として扱い，-10[N] から $+10$[N] までの範囲から選択する．状態メンバーシップ関数の初期値は，(a) の Q-Learning を適用する際の状態空間の分割を参照して設定する．具体的には，ガウス基底関数の中央値と (a) の状態分割（表 5.2）の区間の中央値を一致させるなどした．

(d) ファジィ内挿型 Q-Learning 方法 2（連続値入力/離散値出力，FQL-D）：

行動は上述の 3 値の中から選択し，それぞれの行動ごとにファジィ推論システムを用意する（2.5 節参照）．メンバーシップ関数の初期値は FQL-C と同様である．行動のメンバーシップ関数は 3 つ用意する．

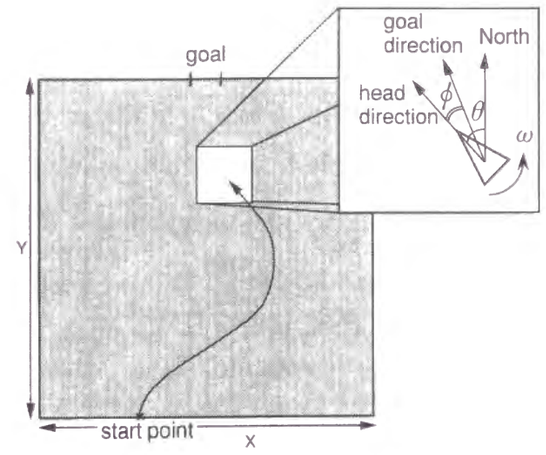


図 5.9: 大型船操舵問題

各方法による学習速度の比較を図 5.8に示す．なお，グラフは 10 回のシミュレーション結果の平均値であり，横軸は試行数，縦軸は成功ステップ数の最大値を表す．この結果からファジィ内挿型 Q-Learning は（FQL-C, FQL-D ともに）従来の Q-Learning や Q-net と比べ，学習速度が速いといえる．これは，ファジィ推論の汎化能力により学習効率が高まった結果であると考えられる．また，離散値行動を扱う場合には，それぞれの行動に対してファジィ推論システムを用いた $OAON$ アーキテクチャにより学習速度が一層向上することが示された．

5.4.2 大型船操舵問題への適用

時間遅れを有する大型船舶の操舵システムの構築を考える．ここでは，特に 1 つのゲートを通過するのが目的であるナビゲーション問題 [71] への適用を試みる．角速度の時定数を T とするとき，大型船舶の操舵システムはつぎの微分方程式で表すことができる．

$$u(t) = T \cdot \frac{d\omega}{dt} + \omega(t) \tag{5.27}$$

$$\theta(t) = \int \omega(t)dt + \theta(0) \tag{5.28}$$

$$\frac{dx}{dt} = V \cdot \sin\theta(t) \tag{5.29}$$

$$\frac{dy}{dt} = V \cdot \cos\theta(t) \tag{5.30}$$

ここで，T: 時定数, V: 速度, u: 操舵量, ω: 旋回角速度, θ: 絶対方位角, x, y: 船の位置座標である．

表 5.3: 大型船操舵問題における状態分割

状態変数	区間	分割数
$\phi[^\circ]$	$[-90,-45]$ $[-45,-10]$ $[-10,+10]$ $[+10,+45]$ $[+45,+90]$	5
$\omega[^\circ]$	$[-35,-10]$ $[-10,+10]$ $[+10,+35]$	3

図 5.9に示すように，船はある一定領域内を移動することができるが，$\pm 90^\circ$ の視野がある．学習エージェントには上記のダイナミクスは与えられておらず，環境からの感覚入力（観測量）として，船首方向とゲートの中心位置との相対角度 ϕ と，船の角速度 ω の 2 つが与えられ，各ステップにおける行動として操舵量 u を決定する．学習の目的はゲートとは反対側にある壁の任意の位置から角度，角速度ともに初期値ゼロの状態から一定の速度で進みゲートに入れるようになることであり，試行に成功すると強化信号（報酬）+1 が学習エージェントに与えられる．操舵中にゲートが $\pm 90^\circ$ の視野の外に出たり，船が領域内から出た場合は失敗試行とし，強化信号（罰）-1 が与えられる．なお，問題に関する各パラメータは，領域の縦 $Y = 500[m]$, 領域の横 $X = 500[m]$, 船の速度 $V = 5[m/s]$, 時定数 $T = 2[s]$, ゲート幅 $Width = 50[m]$ のように設定した．

このような問題設定において，状態空間分割に基づく Q-Learning を適用した場合，Q-net を適用した場合，本提案手法を適用した場合の比較を行う．それぞれの方法は以下のような設定で適用した．なお，学習のパラメータとしては学習率 $\alpha = 0.5$，割引率 $\gamma = 0.99$，温度定数 $T = 0.005$ として学習を行った．

(a) Q-Learning（離散値入力/離散値出力）：
表 5.3に示すように，状態空間を $5 \times 3 = 15$ 個に分割し，操作量として取り得る操舵量 u は $-35, 0, +35[^\circ/s]$ の 3 値とする．

(b) Q-net（連続値入力/離散値出力）：
行動は上述の 3 値の中から選択するものとし，各行動ごとに 8 ユニットの中間層をもつ 3 層の階層型ニューラルネットワークを用いて Q 値を出力する．なお，学習速度の向上のために慣性項の導入などは前節と同様に行った

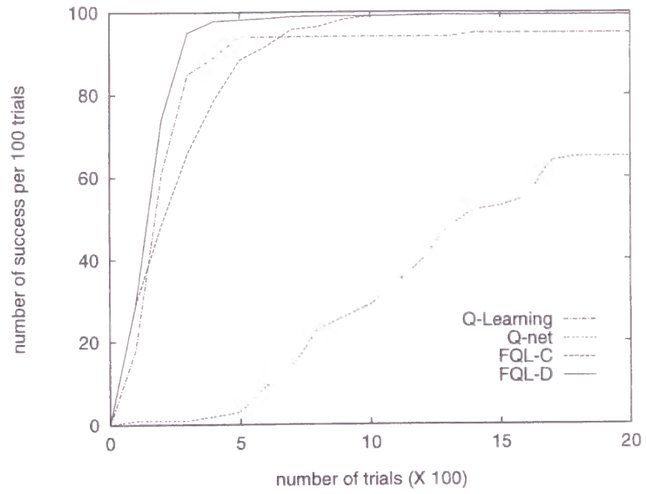


図 5.10: 大型船操舵問題における学習速度の比較

(c) ファジィ内挿型 Q-Learning 方法 1（連続値入力/連続値出力，FQL-C）：
行動を連続値として扱い，$-35[^\circ/s]$ から $+35[^\circ/s]$ までの範囲から選択する．状態メンバーシップ関数の初期値は，状態分割に基づく Q-Learning を適用する際の状態空間の分割（表 5.3）を参照して設定する．

(d) ファジィ内挿型 Q-Learning 方法 2（連続値入力/離散値出力，FQL-D）：
行動は上述の 3 値の中から選択し，各行動ごとにファジィ推論システムを用意する（2.5 節参照）．メンバーシップ関数の初期値は FQL-C と同様である．行動のメンバーシップ関数は 3 つ用意する．

各方法による学習速度の比較を図 5.10に示す．なお，グラフは 10 回のシミュレーション結果の平均値であり，横軸は試行数，縦軸は 100 試行あたりの成功試行数の最大値を表す．この結果からファジィ内挿型 Q-Learning は Q-net と比べて学習速度が速いといえるが，従来の Q-Learning との差はそれほど得られなかった．これは対象問題が時間遅れ系ではあるがここでの問題設定が比較的単純であり，状態分割も良かったことが理由として考えられる．また，離散値行動を扱う場合には，$OAON$ アーキテクチャにより学習能力が若干ではあるが向上することが確認された．

5.4.3 考察

前節では，提案手法のファジィ内挿型 Q-Learning を倒立振子制御問題および大型船操舵問題に適用し，Fig.6 および Fig.8 で示した学習曲線から，状態分割に基づく Q-table を用いた方法や Q-net と比較して，学習速度が優れていることを確認した．Q-table を用いた方法は実現が単純であり，状態分割が良ければ学習効率は決して悪くないと思われるが，それよりも提案手法の方が行動学習が速く進んだ．この理由の 1 つとしては，ファジィ推論のもつ汎化能力が考えられる．

ファジィ内挿型 Q-Learning では，Q 関数はファジィ推論により導出される．したがって，ファジィ推論で用いるファジィルールが学習の過程において更新されるが，1 ステップで更新されるファジィルールは複数個存在する．一方，Q-table ではテーブル内の 1 つの評価値のみが更新される．よって，提案手法では 1 ステップで学習される領域が局所的ではあるがある程度大きく，汎化能力がより強いと考えられる．このため，特に学習の初期において探索による学習結果がその周囲の領域に対して反映されることにより学習効率が高まっていると考えられる．

一方，Q-net は提案手法と同じく関数近似に基づく手法で汎化能力は高いが学習の効率は良くなかった．Q-net においてもある状態に対する評価の変更が行われるとその効果は広く周囲に波及し汎化能力がある．しかし，ニューラルネットワークでは少数のユニット間結合の重みの更新がネットワーク全体に対して影響が及ぶためにかえって，状態の局所的な学習がなかなか進まず，学習効率の悪化を招いていると考えられる．また，ユニット間結合の重みの更新に用いられるバックプロパゲーション (BP) 法も学習回数をかなり必要とすると考えられる．

ここで，この考察を確かめるために，大型船操舵問題においてこれら 3 つの手法に対して，それぞれ 300 試行後（図の左側）と 2000 試行後（図の右側）に学習によって得られる Q 関数を記録し，その推移の様子を図 5.11に示す．まず，Q-table による方法（図 5.11上段）では Q 関数は離散的に与えられるものの，300 試行以降の変動があまりないことから，学習の早い段階から効率的に学習ができているといえる．Q-net による方法（図 5.11中段）では，学習の初期において Q 関数はなかなか学習されていない．ファジィ内挿型 Q-Learning（図 5.11下段）では，ファジィ推論の内挿効果によって複雑な関数が学習の早い段階からなめらかに近似できていると思われる．

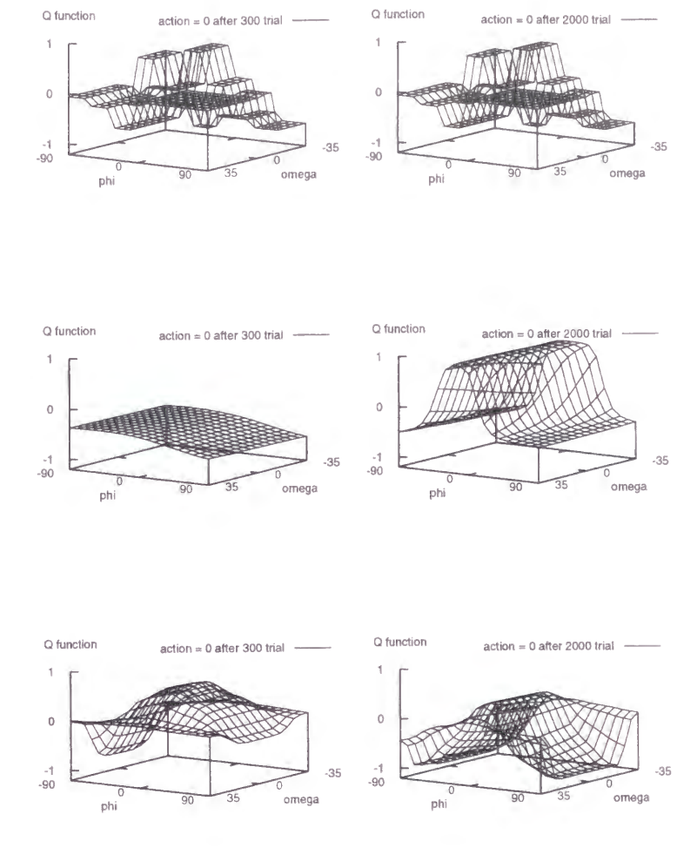


図 5.11: Q 関数の学習の様子 (a) 上段：従来の Q-Learning，(b) 中段：Q-net，(c) 下段：ファジィ内挿型 Q-Learning

5.5 結言

本章では，従来の Q-Learning では困難であった連続値の入力（状態）および出力（行動）を扱うことができるように，Q 値の導出にファジィ推論を導入した新しい学習法であるファジィ内挿型 Q-Learning を提案した．これはファジィルールを用いて行動価値関数（Q 関数）を表現するものであり，Q 関数をなめらかに近似（内挿）することができ，汎化能力が期待できる．また，最急降下法の導入によりファジィルールのパラメータ・チューニングは前件部・後件部ともに学習プログラムが自律的に行うことができる．さらに，行動が離散値の場合には学習の効率化を図る手法として，各行動ごとにファジィ推論システムを用意するアーキテクチャの導入を提案した．なお，本章では示せなかったが，提案手法では人間がもつ先

験的な知識（領域知識）を利用することも容易であり一層の学習の効率化が期待できる．最後に，提案したファジィ内挿型 Q-Learning を倒立振子制御問題と大型船操舵問題の二種類の制御問題に適用し，従来手法と比較することにより，特に学習速度の点においてその有効性を確認した．

しかしながら，提案したファジィ内挿型 Q-Learning にはいくつかの課題および発展の方向性も残されている．まず，ファジィ内挿型 Q-Learning では，ファジィルールのパラメータ・チューニングは前件部・後件部ともに可能であるが，それだけでなくファジィルール自体の追加や削除が学習により実現できればより高い学習効果（精度や汎化能力）が得られると考えられる．また，ファジィ内挿型 Q-Learning の行動選択部では，ある一定数の行動をサンプリングして Q 値を計算してそれらの値を利用して行動を決定しているが，サンプリング数が少ないと行動の選択の幅が狭まり，多いと計算時間が増加するという問題があるため，より良い行動選択手法の導入が望まれる．さらに，学習のより一層の高速化も考えられるが，これについては次章において経験強化を考慮し，Prhofit Sharing 法の考え方を導入した手法を提案する．

本章では，強化学習アルゴリズムとして Q-Learning に着目し，その拡張をいくつかの面から試みたが，本来 Q-Learning はマルコフ性を有する環境に対する手法であり，現実の多くの問題は非マルコフ的であるため，そのような非マルコフ環境における有効な解法を見い出すことが強く望まれる．１つの有望な手法としては確率的傾斜法 [72] などの記憶を用いないアプローチがあり，それに対してファジィ推論を導入することも考えられる．また，別の手法としてはエージェントの経験した履歴を活用する記憶を用いたアプローチがあり，行動パターンのチャンキング [73] などによる履歴情報（エピソード）の効率的な活用が考えられる．

第 6 章
経験強化を考慮した強化学習

6.1 緒言

強化学習は古くから行動心理学の分野で研究されてきたが，工学の分野では初期の人工知能研究における Samuel のチェッカープログラムにまで遡ることができる．また，遺伝的アルゴリズムの分野においては，バケツリレー法 [61] や Profit Sharing 法 [62] などの信頼度伝播アルゴリズムが提案された．これらの手法の多くは，それまでの報酬を獲得した経験を積極的に強化することにより学習途中でもなるべく報酬を獲得し続けるという効率性を重視する経験強化型アプローチとして考えられる．したがって，複雑な問題に対してもある程度効率的に学習がなされるが，未知環境をあまり探索しないために最適性は保証されない．

一方，近年 $TD(\lambda)$ 法 (Temporal Difference 法) [58] や Q-Learning [59] などの動的計画法（Dynamic Programming) に基づいた強化学習アルゴリズムが提案されている．これらの手法は，環境を広く探索することにより結果としてなるべく大きい報酬を得るという最適性を重視した環境同定型アプローチとして捉えることができる．特に，Q-Learning はマルコフ決定問題においてはある条件の下で最適性が保証されているが，複雑な問題に対しては学習に多数の試行錯誤による行動が必要であり学習速度が遅い．

本章では，これら２つのアプローチの長所を活かした学習法として，環境同定型の代表的な手法である Q-Learning に対して，経験強化型の強化学習である Profit Sharing 法（Profit Sharing Plan）の考え方を導入した手法を提案する．この提案する Q-PSP Learning [77] により，学習の高速化・効率化などが期待でき，より現実的な様々な問題に対しても適用できると考えられる．さらに，この提案手法をいくつかの制御問題や自律移動ロボットへ適用することによってその有効性を検証する．

以下，**6.2 節**ではこれまでに提案されている強化学習アルゴリズムについて分類を行なった後，**6.3 節**において，Q-Learning に対して，経験強化型の強化学習である Profit Sharing 法の考え方を導入した手法として Q-PSP Learning を提案する．**6.4 節**では，この提案手法を大型船操舵問題や衝突回避操舵問題，自律移動ロボットなどへ適用することによってそ

の有効性を明らかにする．さらに **6.5** 節では，我々が提案している連続値入出力（状態・行動）を扱うファジィ内挿型 Q-Learning [76] に対して Q-PSP Learning の枠組みを導入した新たな学習法を提案し，**6.6** 節において倒立振子制御問題への適用を行いその有効性を明らかにする．

6.2 強化学習アルゴリズムの分類

本節では，さまざまなものが提案されている強化学習アルゴリズムについて分類を行なう．分類の第一の軸は，環境のクラスである．すなわち，環境のクラスは状態遷移にマルコフ性を仮定するかどうかで大きく分けられる．分類の第二の軸は，接近の指向性である．前章で少し述べたように，結果としてなるべく大きい報酬を得るという**最適性**と，学習途中でもなるべく報酬を獲得し続けるという**効率性**の二つの側面がある．最適性については，環境を広く探索することにより得られるので，最適性を重視する接近法は環境同定型 (Exploration-Oriented) のアプローチと呼ばれる．また，効率性は，それまでの報酬を獲得した経験を積極的に強化することにより得られるため，効率性を重視する接近は経験強化型 (Exploitation-Oriented) のアプローチと呼ばれる．

6.2.1 環境同定型アプローチ

環境同定型アプローチとしては，$TD(\lambda)$ 法，Q-Learning, k-確実探査法などが提案されている．$TD(\lambda)$ 法，Q-Learning については前章で述べたので，ここでは極端に環境同定に徹した方法である k-確実探査法を簡単に紹介する．

k-確実探査法は，効率の良い環境同定を実現するために行動選択において，各ルールの選択回数のばらつきを抑えながら，全てのルールを最低 k 回選択することを実現する．ここで，選択回数が k 回以上になっているルールを **k-確実**という．k-確実探査法では，行動選択後の状態遷移確率および得られる報酬の期待値を最尤推定により同定し，k が十分大きくなって統計的推定が正しくなったとき，最適な行動が保証される．

6.2.2 経験強化型アプローチ

経験強化型アプローチとしては，Samuel の Checker Player や分類子システム (Classifier System) で用いられるバケツリレー法, Profit Sharing 法などがある．

Samuel のチェッカープログラム Checker Player は，ゲームの木の探索に用いる評価関数をゲームの勝敗を報酬として強化する．二人のエージェントが互いにゲームをしながら学習するので，学習するマルチエージェント系という興味深いクラスを扱っていた．

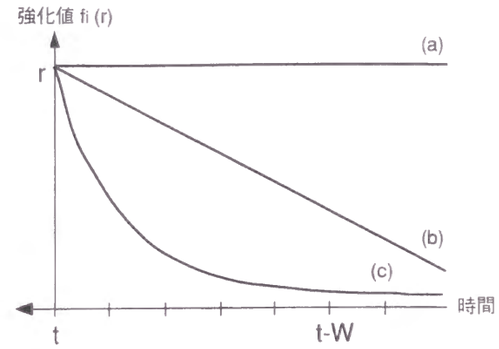


図 6.1: 強化関数の例

Profit Sharing 法については，合理性に関する定理が得られており，以下に簡単に紹介する．Profit Sharing 法において，報酬までのステップ数と報酬の分配率を対応づける関数を**強化関数**という．この強化関数については，明らかに無駄なルール（無効ルール）を強化しないという局所的な合理性と，必ずいくらかの報酬を継続して獲得するという大局的な合理性を満足する条件が以下のようになる．

$$\sum_{j=N}^{W} f_j < f_{N-1} \tag{6.1}$$

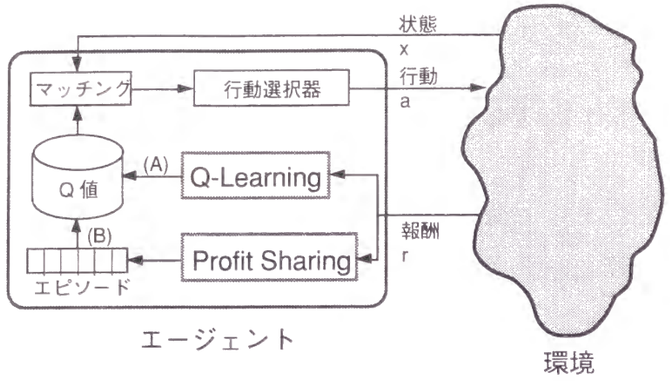
式 (6.1) を満たす最も簡単な強化関数は，等比減少関数が考えられる（図 6.1(c)）．従来用いられていた定数関数（図 6.1(a)）や等差減少関数（図 6.1(b)）は，この定理を満たさず，無駄なルールが強化されることがある．

6.3 経験強化を考慮した Q-Learning の提案

6.3.1 Q-PSP Learning の枠組み

前章で述べたように，1-Step Q-Learning では，得られた報酬の伝搬はバケツリレーアルゴリズム (BBA) と同じく，次の実行の際に 1 段階伝搬するだけであるため，学習速度は遅い．これに対して Profit Sharing 法 (PSP) では，実行されたルールの履歴をエピソードとして記憶しておき，報酬が得られるたびに過去に実行されたルールの強さを報酬の値を減じながら一括して変更するため，学習の進行は比較的速い．そこで本章では，Q-Learning の高速化・効率化を図るために，1-Step Q-Learning に Profit Sharing 法を導入した手法（Q-PSP Learning）を提案する（図 6.2参照）．

Q-PSP Learning では，Q-Learning と同様に各ルールが Q 値をもっており，各ステップ



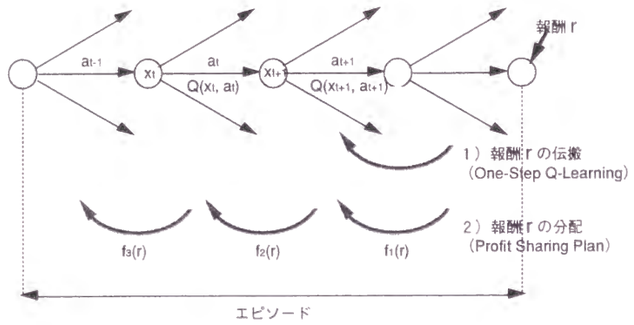


図 6.2: Q-PSP Learning の枠組みと報酬の伝播

において Q 値の更新が 1-Step Q-Learning によりなされるとともに，実行したルール系列をエピソードとして記憶しておき，ゼロでない報酬が得られた際に報酬を一括して過去に実行されたルールまで伝搬させる．これにより，Q-Learning における報酬の伝搬を速めることができ，学習速度の向上が期待できる．

ただし，Q-Learning では，マルコフ問題においては学習率 α を適切に減少させると Q 値が収束し，そのとき各状態において最大の Q 値をもつルールの選択が最適政策となることが証明されている [59]．つまり，Q-Learning は結果としてなるべく大きい報酬を得るという最適性を重視したアルゴリズムであるといえるのに対して，Q-PSP Learning では，収束性や最適性に関する考察はまだされていない．

6.3.2 Q-PSP Learning の手順

まず，表 6.1に，提案する Q-PSP Learning のアルゴリズムの概略を示す．

時刻 t において状態が x_t であって行動 a_t を実行した結果，状態は x_{t+1} に遷移し，報酬 r

表 6.1: Q-PSP Learning のアルゴリズム

(1) 現在の状態 x_t を観測する

(2) Q-Table により Q 値を求める

(3) Q 値に基づいて行動を選択する

(4) 選択された行動 a_t を実行し次状態 x_{t+1} に遷移するとともに実行したルールをエピソードに記憶する

(5) Q 値の更新幅 $\Delta Q(x_t, a_t)$ を計算し，Q 値の更新を行う

(6) 報酬 $r = 0$ の場合，Step (1) にもどる

報酬 $r \neq 0$ の場合，Profit Sharing 法による Q 関数の変更を行う

が得られたとき，Q 値は 1-Step Q-Learning と同じつぎの更新式で変更される．

$$Q(x_t, a_t) \leftarrow (1-\alpha)Q(x_t, a_t) + \alpha(r + \gamma \max_b Q(x_{t+1}, b)) \tag{6.2}$$

また，実行したルール系列をエピソードとして記憶しておき，ゼロでない報酬が得られた際に報酬を一括して過去に実行されたルールまで伝搬させる．すなわち，エピソードに参加したルール R_i の Q 値は以下の式に従って更新される．

$$Q(x_t, a_t) \leftarrow (1-\alpha')Q(x_t, a_t) + \alpha' f_i(r) \tag{6.3}$$

ただし，$\alpha'(0 < \alpha' < 1)$ は学習率である．また，$f_i(r)$ は強化関数であり，エピソードの最後から数えて i ステップ前のルールに分配する報酬の大きさを決定する．この強化関数 $f_i(r)$ については，文献 [83] に示された強化関数に関する合理性を保証する条件を考慮し，$f_i(r) = r(\gamma')^i$ とする $(0 < \gamma' < 1)$．

ここで，エピソードの記憶には，記憶容量が有限（最大記憶ルール数を N 個とする）の記憶（メモリ）を用い，記憶容量を越えてルール系列が入ってくる場合には，古いものから順に忘却してゆくメカニズムを採用する．これにより，最新の N 個のルール系列をエピソードとして保持することができる．

また，今回のシステムでは 1-Step Q-Learning における学習率 α は一定としたが，PSP における学習率 α' は $\alpha' = \alpha$ から学習の進行とともに徐々に小さくしてゆく．これは，Q 値の更新が過度になされるのを抑えるためである．

行動決定の方法としては，次式に示す Boltzmann 分布に基づく確率的な行動選択をここ

では用いる．

$$P(a|x) = \exp(Q(x,a)/T)/\sum_{b\in A}\exp(Q(x,b)/T)$$

ただし，T は温度定数であり，この値が大きいほど行動はよりランダムになり積極的な探索を行う．

6.3.3 Q-PSP Learning の特徴

先にも述べたように，Q-PSP Learning では，Q 値の更新が 1-Step Q-Learning だけでなく Profit Sharing 法 によっても行われる．すなわち，各ステップにおいて Q 値の更新が 1-Step Q-Learning によりなされるとともに，学習エージェントが実行したルール系列をエピソードとして記憶しておき，正の報酬を受け取った際（行動系列が成功した場合）にはエピソード内のルールの Q 値を一括して増加させ，負の報酬（罰）を受け取った際（行動系列が失敗に終わった場合）にはエピソード内のルールの Q 値を一度に減少させる．したがって，Q-PSP Learning は，成功事例と失敗事例の両方の経験を積極的に強化するため，経験強化型の強化学習法の側面をもつ．

それに対して，Q-Learning は結果としてなるべく大きい報酬を得るという最適性を重視した環境同定型のアルゴリズムであり，ある条件のもとでは最適性（収束性）が証明されている．しかしながら，Q-PSP Learning では利益共有法（PSP）の経験強化の側面を取り入れた学習法であるため，最適性は一般に保証されない．しかし，実際的な大規模問題に対しては，最適性よりはむしろ経験強化に重点を置いた学習法が有効な場合が多い．この環境同定 (Exploration) と経験強化 (Exploitation) のバランスを如何にとるかは重要な問題 [80] であり，Q-PSP Learning では，このバランスを考慮することができる手法といえる．すなわち，Q-PSP Learning では，最大記憶ルール数 N や学習率 α' を問題に応じて変更することができ，$N=0$ あるいは $\alpha'=0$ の場合は，Q-Learning に等しくなる．

6.4 例題への適用

6.4.1 大型船操舵問題

本節では，時間遅れを有する大型船の操舵システムの構築を考える．ここでは，とくに 1 つのゲートを通過するのが目的であるナビゲーション問題 [71] への適用を試みる．角速度の時定数を T とするとき，大型船の操舵システムはつぎの微分方程式で表すことができる．

$$u(t) = T\cdot\frac{d\omega}{dt} + \omega(t) \tag{6.4}$$

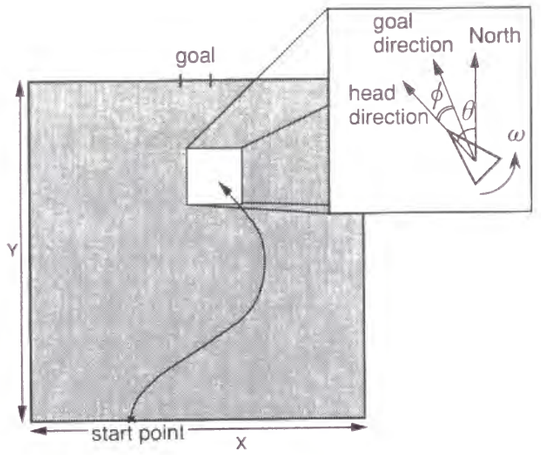


図 6.3: 大型船操舵問題

$$\theta(t) = \int \omega(t)dt + \theta(0) \tag{6.5}$$

$$\frac{dx}{dt} = V\cdot\sin\theta(t) \tag{6.6}$$

$$\frac{dy}{dt} = V\cdot\cos\theta(t) \tag{6.7}$$

ここで，T: 時定数, V: 速度, u: 操舵量, ω: 旋回角速度, θ: 絶対方位角, x, y: 船の位置座標である．

図 6.3に示すように，船はある一定領域内を移動することができるが，$\pm 90°$ の視野がある．学習エージェントには上記のダイナミクスは与えられておらず，環境からの感覚入力（観測量）として，船首方向とゲートの中心位置との相対角度 ϕ と，船の角速度 ω の 2 つが与えられ，各ステップにおける行動として操舵量 u を決定する．学習の目的はゲートとは反対側にある壁の任意の位置から角度，角速度ともに初期値ゼロの状態から一定の速度で進みゲートに入れるようになることであり，ゴールに入るか，壁にぶつかるか，ゴールを見失うかにより停止するまでを 1 試行 (trial) とみなす．試行に成功すると強化信号（報酬）$+1$ が学習エージェントに与えられるとともに，操舵中にゲートが $\pm 90°$ の視野の外に出たり，船が領域内から出た場合は失敗試行とし，強化信号（罰）-1 が与えられる．なお，領域の横幅 $X = 500\ (m)$，領域の縦幅 $Y = 360\ (m)$，船の速度 $V = 3\ (m/s)$，時定数 $T = 3\ (s)$，ゴール幅 $Width = 50\ (m)$ としてシミュレーションを行った．なお，操舵量は 0.2 秒毎に加えるものとする．

ここでは，環境からの入力データとして，ゴールと船首方向の相対角度 ϕ と船の角速度

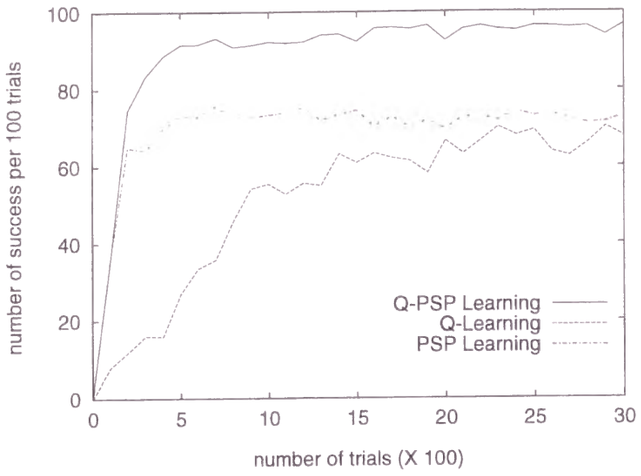


図 6.4: 大型船操舵問題における学習速度の比較

ω の 2 変数を用い，これらの 2 次元の状態空間の分割を行う．この 2 変数に関して，範囲 $-90° \leq \phi \leq 90°, -35°/s \leq \omega \leq 35°/s$ の状態空間を $5 \times 7 = 35$ 個の部分状態空間に分割する．操作量としては，操舵量 $u = -35, -17.5, 0, +17.5, +35$ $(°/s)$ の 5 値とする．すなわち，各部分状態空間においてこれら 5 値のうちのいずれを出力すべきかを学習させる．なお，学習のパラメータとしては，$T = 0.02, \alpha = 0.4, \gamma = 0.95, N = 20, \alpha' = 0.4, \gamma' = 0.9$ などとした．

シミュレーション結果として，100 試行毎の成功回数（10 セットの平均）の変化を図 6.4 に示す．この図から Q-PSP Learning は Q-Learning よりも学習が比較的早く進行していることが分かる．

6.4.2 衝突回避操舵問題

本節では，より大規模な問題として，動的な障害物である領域侵入船が存在するような状況下での大型船操舵問題（図 6.5参照）への適用を考える．大型船の操舵システムは，前節と同じ微分方程式でモデル化できる．

ここでは，環境からの入力データとして，自船進行方向とゴールの中心位置との相対角度 ϕ，自船の旋回角速度 ω，侵入船との距離 D，侵入船との相対角度 δ，侵入船との相対角速度 η の 5 変数を用いる．船を制御する操舵量は，離散値で 5 値とする．

また，ここでの学習目標は，ゴールとは反対の壁の任意の位置から出発した操舵船が一定の速度 V で進み，速度 V' で進む侵入船に衝突しないように回避してうまくゴールに入るよ

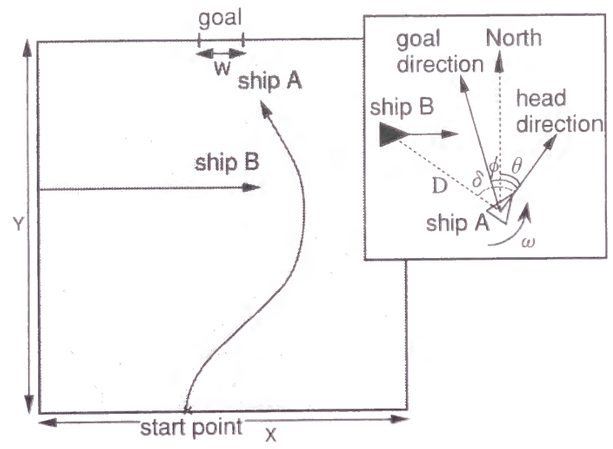


図 6.5: 衝突回避操舵問題

うになることである．操舵船が侵入船を回避しゴールに入れば強化信号（報酬）+1 が与えられ，侵入船に衝突するか，壁にぶつかるか，あるいはゴールを見失うかすれば，失敗とみなし強化信号（罰）−1 が与えられる．

なお，問題に関する各パラメータは，領域の横幅 $X = 500$ (m)，領域の縦幅 $Y = 360$ (m)，自船の速度 $V = 3$ (m/s)，侵入船の速度 $V' = 3$ (m/s)，時定数 $T = T' = 2$ (s)，ゴール幅 $Width = 50$ (m) としてシミュレーションを行う．なお，操舵量は 0.2 秒毎に加えるものとする．また，今回のシミュレーションでは簡単のため，侵入船は左右いずれかかから一定速度 V' で水平方向にだけ進むものとする．

環境からの入力（観測変量）は，$(\phi, \omega, D, \delta, \eta)$ の 5 変数であり，これらの 5 次元の状態空間を $5 \times 7 \times 2 \times 5 \times 5 = 1750$ 個の部分状態空間に分割する．操作量としては，操舵量 $u = -35, -17.5, 0, +17.5, +35$ $(°/s)$ の 5 値とする．すなわち，各部分状態空間においてこれら 5 値のうちのいずれを出力すべきかを学習させる．なお，学習パラメータとしては，$T = 0.02, \alpha = 0.3, \gamma = 0.95, N = 20, \alpha' = 0.3, \gamma' = 0.8$ などとした．

シミュレーション結果として，100 試行毎の成功回数（10 セットの平均）の変化を図 6.6 に示す．ここでは，比較のため，Q-Learning を適用した場合と PSP だけを適用した場合の学習曲線も描いている．この図から分かるように，ここで取り上げたような大規模な問題においては，Q-PSP Learning は Q-Learning よりも学習速度が極めて高速であると言える．また，Profit Sharing 法（PSP）と比較した場合，学習の初期段階での立ち上がりは同程度であるが，それ以降の伸びが Q-PSP Learning の方が良いと言える．この理由として

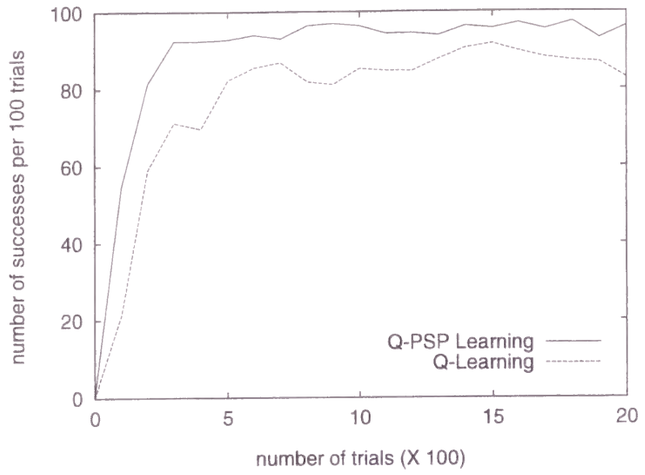


図 6.6: 衝突回避操舵問題における学習速度の比較

は，PSP が過去の良い経験のみを重視しているのに対して，Q-PSP Learning では良い経験を保持しつつ，さらに良い行動を見つけるための探索 (Exploration) をしていることが考えられる．

6.4.3 自律移動ロボットの行動形成問題

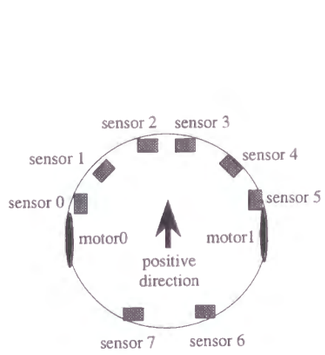
本節では，前章で述べた強化学習法を自律移動ロボットの行動形成に応用する．実ロボットに強化学習を応用する場合，シミュレーションと同じ条件で学習させることは学習時間やロボットの物理的な制約などから困難である．本節では，実環境に即して問題を設定し，ロボットに目的行動を獲得させることを試みる．

問題設定と学習方法

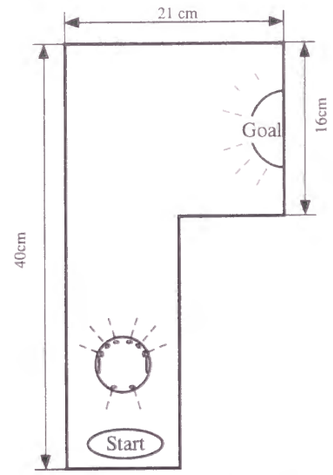
実験には，8 個ずつの近接センサ (proximity sensor) と光センサを備えている小型移動ロボット *Khepera* [74]（図 6.7(a) 参照）を用いる．ロボットは図 6.7(b) に示される環境の中を移動する．ロボットに与えられたタスクは，周囲の壁にぶつからないように通路を進んで右に曲がってゴールである光源に到達することである．また，ロボットが選択可能な行動は，直進，右旋回，左旋回の 3 つとする．与えられる強化信号は，壁に衝突したときは罰として -1.0，一回も壁に衝突しないでゴールに到達したときは報酬 $+3.0$，壁に衝突しながらもゴールに到達したときは報酬 $+1.5$ を与えることとする．

本研究においては状態空間を構成するために用いるセンサは，図 6.7(a) に示すようにセンサ番号 0 から 6 までの近接センサとセンサ番号 1,2,4 の光センサとする．各センサそれぞれについて適当なしきい値を設定し，得られるセンサ値が分割されたどの区間に属するかで各センサにおける観測情報を離散化し，その組み合わせによって状態を定義する．具体的には，各センサの区間の分割方法を表 6.2のように定義する．

壁に衝突してでもスタートしてからある一定のステップ数 $STEP$ 以内でゴールに到達した試行を成功試行，$STEP$ 以上のステップ数かかってもゴールに到達できなかった試行を



(a) 小型移動ロボット "Khepera" とそのセンサ配置

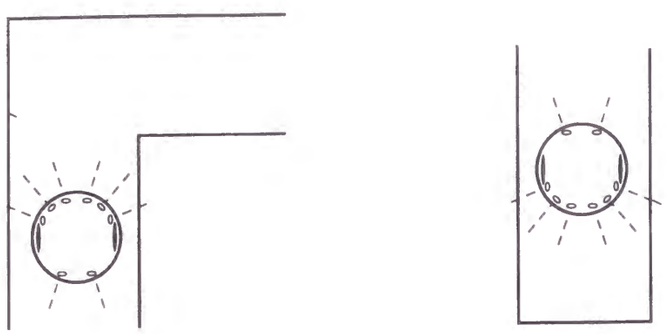


(b) ナビゲーション環境

図 6.7: 移動ロボット *"Khepera"*とタスク環境

表 6.2: センサ情報に基づく状態分割

センサ No.	区間	分割数
近接センサ 0,5	$[0, 300]$ $[300, 700]$ $[700, 1000]$	3
近接センサ 1,4,6	$[0, 300]$ $[300, 1000]$	2
近接センサ 2,3	$[0, 300]$ $[300, 500]$ $[500, 700]$ $[700, 1000]$	4
光センサ 1,2,4	$[400, 520]$ $[150, 400]$	2



(a) (b)

図 6.8: PerceptualAliasing の発生

失敗試行としてこれを 1 試行とする．失敗試行，成功試行以外は壁に衝突しながらもゴールに到達した試行であり，多くの試行はこの試行である．

試行が失敗したときは，そこでその試行をやめ，再びスタート地点に戻して学習をやり直す．このようなことを行なうのは，Perceptual Aliasing（知覚の同一視）[75] による学習への悪影響を少しでも減らすためである．

この環境における移動ロボットの行動形成において,Perceptual Aliasing は以下のような場合において生じる．ロボットが図 6.8(a) のように左右両側に壁を感知している状況を考える．一方，図 6.8(b) のような状況においてもロボットは左右両側に壁を感知している．現在得られるセンサ情報のみを用いて状態を定義しているかぎりロボットはどちらも同じ状態にある．しかし (a) のような状況においては直進すればコーナー方向に進むことができるのに対し,(b) においては直進すればすぐに前方の壁を感知してしまう．このように状態は同一であるのに取るべき行動が異なるような状態が Perceptual Aliasing である．このような場合においては, その時点での状態のみから最適な行動が決められるのではなく, その状態へ遷移するに至った以前の状態にも依存して行動を決定しなければならないということになる．こうした隠れマルコフ状態の下では Q 値は最適な値に収束せず, 学習能力を著しく低下させる．

こうした隠れマルコフ状態を回避するには, 状態の定義を変えたり, 別の種類の情報を用いることによって隠れマルコフ状態かどうかを判別するなどということが考えられる．しかし, 本研究においては, このような隠れマルコフ状態の回避を行なっていないのでこのような措置をとることにする．

表 6.3: 実ロボットにおける失敗試行数と成功試行数の変化

	No. of failure trials		No. of success trials	
trial	QL	Q-PSP	QL	Q-PSP
~20	5	4	0	1
~40	5	1	0	5
~60	2	5	0	2
~80	1	1	0	6
~100	4	1	1	9

シミュレーション実験での結果

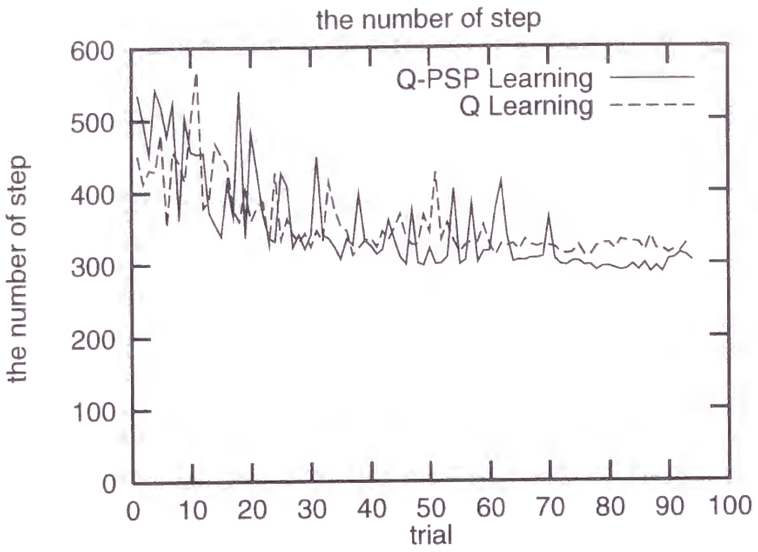
Q Learning と Q-PSP Learning の両手法におけるゴールに到達するまでのステップ数，壁への衝突回数（10 試行セットを平均したもの）を図 6.9に示す．

これらの結果からシミュレーション上においてはパフォーマンスにあまり大きな差がみられないことがわかる．これは N と γ' の大きさに関係していると考えられる．スタートしてからゴールに到達するまでに平均して約 160 回前後の状態遷移をしており,$N = 5$ では報酬の分配サイズが小さく, 効果が小さいと考えられる．また，ゴールに到達する経験をしなければひたすら失敗を繰り返すという経験強化型の強化学習の特徴を表した試行も見られた．

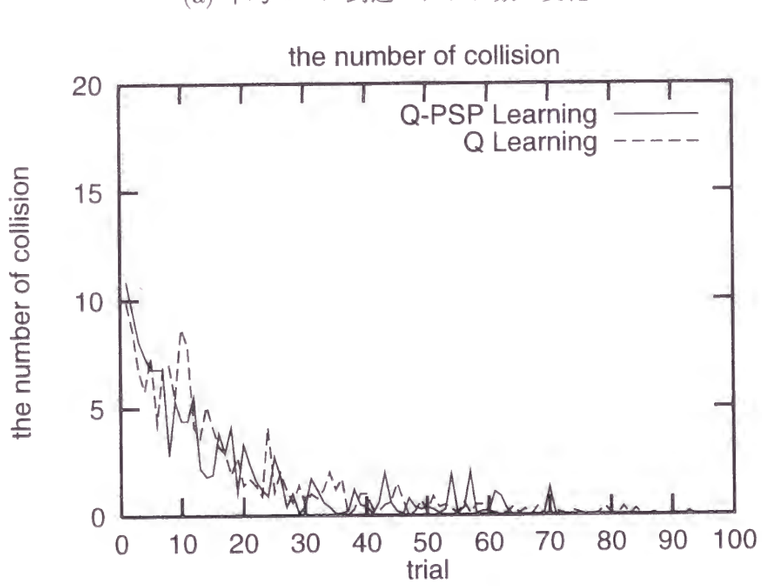
実ロボットにおける学習結果

表 6.3に Q-Learning と Q-PSP Learning の 20 試行ごとの失敗試行数と成功試行数の変化を示す．また失敗試行を除いた試行の Q-Learning と Q-PSP Learning それぞれについてのゴール到達までのステップ数，衝突回数をプロットしたものを図 6.10に示す．Q-PSP Learning において，100 試行行なうのにかかるステップ数は 23027 ステップであり，Q-Learning と比べてかなり少ない．

Q-PSP Learning においては，はじめは Q-Learning と同様，ランダムな探索に基づいて行動し，図 6.11(a),(b) に示すような経路をたどることが多い（40 試行位まで）．しかしその間にもゴールに到達する，壁に衝突するとその経験が Q-Learning の場合よりもより積極的に評価され強化される．そして曲り角においてまっすぐゴールに向かうために右旋回をとるという有効なルールがいち早くそしてより強く強化される．そのため試行を重ねるにつれ Q-Learning の場合と比べて成功試行の数は多くなる．Q-PSP Learning を用いたときの形成されたロボットの軌跡は，図 6.11(c) に示すように，狭い通路をロボットが完全に抜け

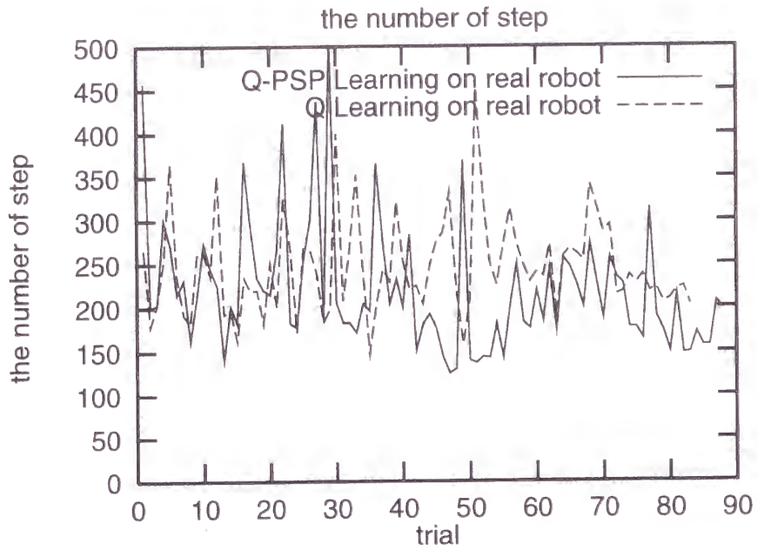


(a) 平均ゴール到達ステップ数の変化

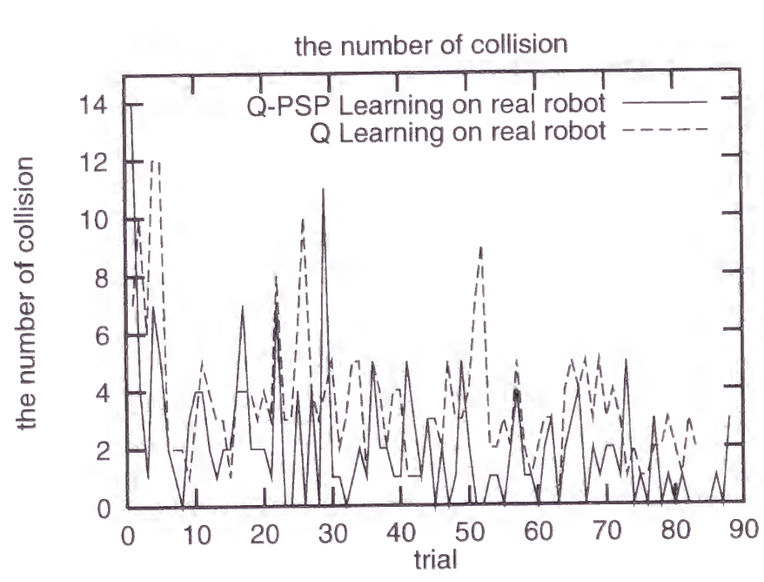


(b) 平均衝突回数の変化

図 6.9: シミュレーション実験における学習結果



(a) 平均ゴール到達ステップ数の変化



(b) 平均衝突回数の変化

図 6.10: 実ロボットにおける学習結果

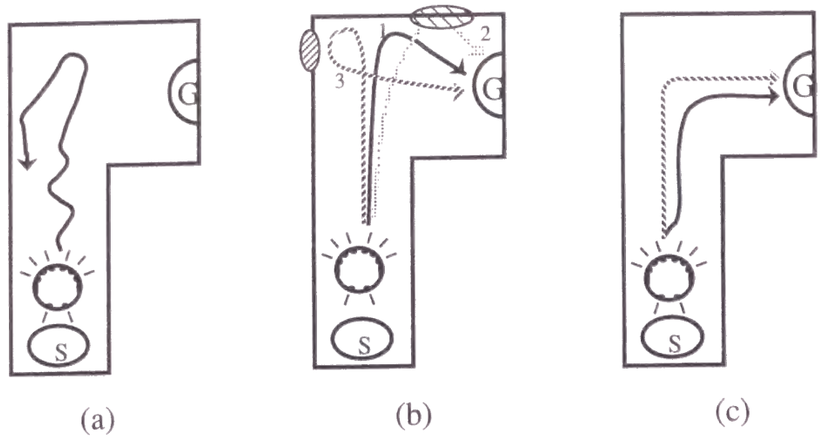


図 6.11: 実ロボットにおいて獲得された行動パターン

きった後に右旋回の行動を取り続けてゴール方向にまっすぐ向いてから直進を続けてゴールに到達するようになる.

シミュレーションにおける実験と比べて実ロボットでの実験の方が Q-PSP Learning の有効性が顕著に現れたのは,N のサイズの影響によると考えられる. 実環境においては, スタートしてからゴールに到達するまでに平均して状態遷移の回数は約 70 回前後であり,N のサイズがシミュレーションの場合に比べて相対的に大きい. このため報酬がより早くスタート地点近くまで伝播されていると考えられる.

Q-PSP Learning を導入することによってうまくゴールに到達する経路が見つかればその経験が強化されて最適ではないにしても壁にぶつからないでゴールに到達する目的行動を獲得することができる. しかし第 77 試行においてみられるように最適行動の選択に失敗しその経路をはずれれば大きく失敗してしまうことが予想される.

6.4.4 考察

これらの例題において, Q-PSP Learning は Q-Learning に比べ学習の速度が極めて速い上に, 学習の結果も優れていることが示されている. 特に, 状態数が極めて多くなる衝突回避操舵問題 (4.2 節) では, Q-PSP Learning の有効性が顕著に示された. また, 各例題においては, それぞれ有効な制御則が学習により獲得された. 特に, 衝突回避操舵問題では, 学習終了後において, 侵入船との衝突をうまく回避しながらゴールに向かう操船がなされるようになった. 本章で取り上げた 2 つの例題だけでなくさまざまな問題に対して Q-PSP

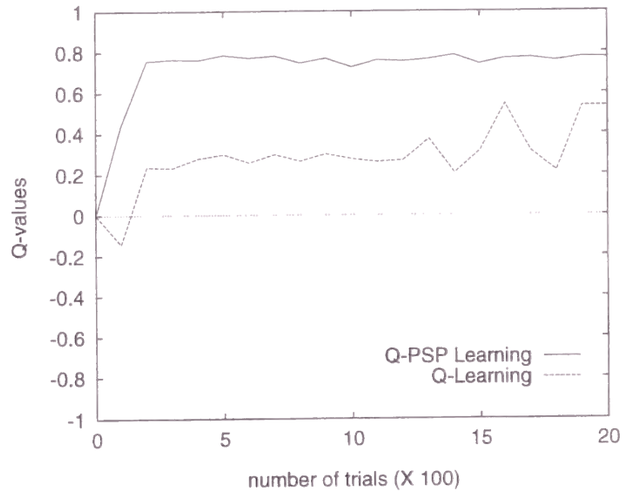


図 6.12: 大型船操舵問題における有効なルールの Q 値の推移

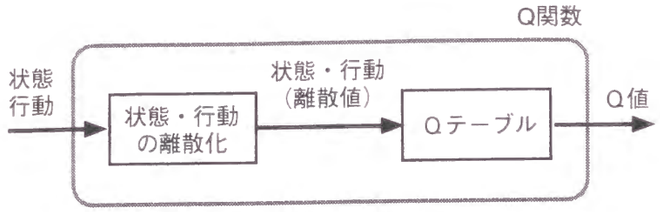
Learning の適用を試みているが, 現実的な大規模の問題になるほど Q-PSP Learning の方が Q-Learning よりも有効であると思われる.

Q-PSP Learning が比較的うまくいった要因としては, Q-Learning よりも報酬の伝搬が速められ一部の有効なルールがいち早く強化されるからだと考えられる. 一例として, 図 6.12に, 大型船操舵問題 (6.4.1 節) において有効と考えられるルール群中の一つのルール (if $15° \leq \phi \leq 45°$, $-8.75°/s \leq \omega \leq 8.75°/s$ then $u = -35°/s$) に関する Q 値の時間的推移を示す. このように, Q-PSP Learning では Q-Learning に比べて有効なルールがいち早くそしてより強く強化されていることがわかる.

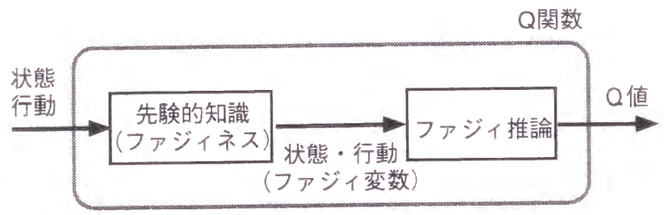
6.5 Profit Sharing 法を導入したファジィ内挿型 Q-Learning

6.5.1 提案手法の枠組み

我々は, 従来の Q-Learning では困難であった連続値の入出力 (状態・行動) を扱うことを可能にするために, Q 値の導出にファジィ推論を導入した新しい学習法であるファジィ内挿型 Q-Learning [76] を提案している (図**??**参照). これはファジィルールを用いて行動価値関数 (Q 関数) を表現するものであり, Q 関数をなめらかに近似 (内挿) することができ, 汎化能力が期待できる. また, 最急降下法の導入により Q 関数を表すファジィルールのパラメータ・チューニングは前件部・後件部ともに学習プログラムが自律的に行うことができるのも特徴である.



(a) Q-Tableを用いたQ関数（従来手法）



(b) ファジィ推論法を用いたQ関数（提案手法）

図 6.13: 従来手法と提案手法の枠組み

本章では，ファジィ内挿型 Q-Learning の学習の高速化を図るために，6.3 節で提案した Q-PSP Learning の枠組みに基づき，ファジィ内挿型 Q-Learning に Profit Sharing 法を導入した手法を提案する．この手法では，ファジィ内挿型 Q-Learning の各ステップにおいて Q 値の更新がなされるとともに，実行したルール系列をエピソードとして記憶しておき，ゼロでない報酬が得られた際に報酬を一括して過去に実行されたルールまで伝搬させる手法である Profit Sharing 法による Q 関数の更新を行う．

したがって，提案手法では，従来の Q-Learning による更新（Step (6)）と Profit Sharing 法による更新（Step (8)）の 2 通りにより Q 関数の更新を行うために，ファジィ内挿型 Q-Learning の手順の一部（Step (4), Step (8)）を変更する必要があり，全体の流れを表 6.4に示す．

6.5.2 エピソードの記憶

ファジィ内挿型 Q-Learning ではファジィ推論を用いているため，各ステップで発火（実行）されるファジィルールは一般に複数あり，それら各ルールの推論結果を適合度を考慮して統合（非ファジィ化）することにより Q 値が導出される．したがって，Step (4) において各ステップで発火した複数のルールとそれらの適合度をエピソードとして記憶すべきであるが，その代わりに各ステップでの状態・行動（連続量）の履歴を記憶することにより，エピソードから各ステップでの状態と行動を取り出して発火したルールとその適合度を求めることが可能となる．

なお，エピソードの記憶には，記憶容量が有限（最大記憶ルール数を N 個とする）の記憶（メモリ）を用い，記憶容量を越えてルール系列が入ってくる場合には，古いものから順に忘却してゆくメカニズムを採用する．

6.5.3 Q 関数の更新

Step (6) における Q 関数の更新では，従来の Q-Learning と同じつぎの計算式で Q 値の更新幅 ΔQ が求められる．

$$\Delta Q(x_t, a_t) = \alpha(r_t + \gamma \max_b Q(x_{t+1}, b) - Q(x_t, a_t)) \tag{6.8}$$

また，Step (8) における Profit Sharing 法による Q 関数の更新では，あらかじめ実行したルール系列をエピソードとして記憶しておき（Step (4)），ゼロでない報酬が得られた際に報酬を一括して過去に実行されたルールまで伝搬させる．すなわち，エピソードに参加したルール R_i の Q 値の更新幅 $\Delta Q'$ は以下の式に従って計算される．

$$\Delta Q'(x_t, a_t) = \alpha'(f_i(r) - Q(x_t, a_t)) \tag{6.9}$$

ただし，α' は学習率 $(0 < \alpha' < 1)$ である．また，$f_i(r)$ は強化関数であり，エピソードの最後から数えて i ステップ前のルールに分配する報酬の大きさを決定する．

提案手法では，これらの $\Delta Q, \Delta Q'$ をもとに Q 関数を表すファジィルールの更新を行う．具体的には，最急降下法を用いることによりこれらの二乗誤差を最小化する方向にファジィルールの後件部パラメータが更新される．

6.6 倒立振子制御問題への適用

倒立振子制御問題では，振子の角度 θ, 角速度 $\dot{\theta}$, 台車の位置 x, 速度 $\dot{x}$ を観測量とし，行動として台車に加える水平外力（操作量）F を決定する．倒立振子系を表す微分方程式はつぎのように記述することができる [70].

$$\begin{aligned}
\ddot{\theta}_t &= \frac{g \sin\theta_t + \cos\theta_t \left[\frac{-F_t - ml\dot{\theta}_t^2 \sin\theta_t + \mu_c \mathrm{sgn}(\dot{x}_t)}{m_c + m}\right] - \frac{\mu_p \dot{\theta}_t}{ml}}{l\left[\frac{4}{3} - \frac{m\cos^2\theta}{m_c + m}\right]} \\
\ddot{x}_t &= \frac{F_t + ml\left[\dot{\theta}_t^2 \sin\theta_t - \ddot{\theta}_t \cos\theta_t\right] - \mu_c \mathrm{sgn}(\dot{x}_t)}{m_c + m}
\end{aligned}$$

表 6.4: Profit Sharing 法を導入したファジィ内挿型 Q-Learning のアルゴリズム

(1) 現在の状態 x_t を観測する

(2) ファジィ推論により Q 値を求める

(3) Q 値に基づいて行動を選択する

(4) 選択された行動 a_t を実行し次状態 x_{t+1} に遷移するとともに実行したルールをエピソードに記憶する

(5) ファジィ推論により Q 値を求める

(6) Q 値の更新幅 $\Delta Q(x_t, a_t)$ を計算する

(7) Q 関数（ファジィルール）の更新を行う

(8) 報酬 $r = 0$ の場合，Step (1) にもどる

報酬 $r \neq 0$ の場合，Profit Sharing 法による Q 関数の変更を行う

本研究では，Runge-Kutta-Gill 法を用いて計算機シミュレーションを行う．また問題に関する各パラメータは重力加速度 $g : 9.8[m/s^2]$, 台車の質量 $m_c : 1.0[kg]$，振子の質量 $m : 0.1[kg]$, 振子の半分の長さ $l : 0.5[m]$, 台車とレールの摩擦係数 $\mu_c : 0.0005$, 振子とヒンジの摩擦係数 $\mu_p : 0.000002$ などとした．

学習エージェントには上記のダイナミクスは与えられておらず，振子が 12 度以上傾いたとき，あるいは台車がレールからはみ出したとき失敗したとみなし，強化信号（罰）-1 が与えられるものとする．10,000 ステップ以上失敗なかったときには成功したものとし，つぎの試行を行う．

このような問題設定の下で，ファジィ内挿型 Q-Learning と Profit Sharing 法を導入した提案手法の 2 つの手法についてシミュレーションを行い，それらの学習速度の結果（10 セットの平均）を図 6.14に示す．この図より，若干ではあるが学習の高速化が図られていることが分かるが，期待していたほどの向上は得られなかった．その理由としては，メンバーシップ関数が重なっているところでの Q 値の更新は複数のファジィルールに影響を与えるために，あまり学習率を高くできないことが考えられる．

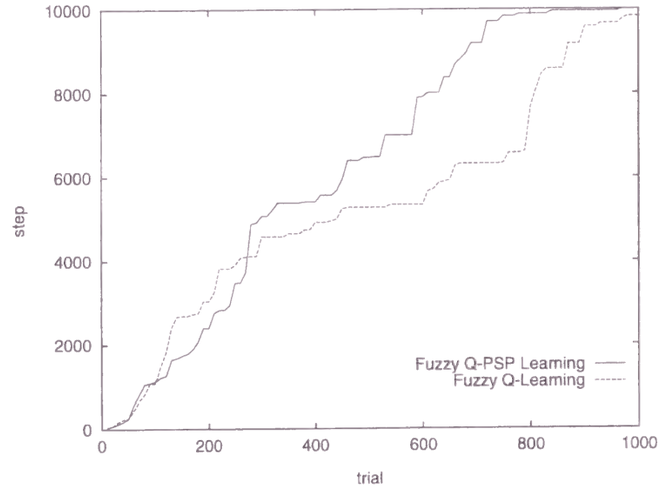


図 6.14: 倒立振子制御問題における学習速度の比較

6.7 結言

本章では，強化学習の代表的な手法の一つである Q-Learning に対して，学習の高速化・効率化を図るために，分類子システムで用いられる経験強化型の強化学習アルゴリズムである Profit Sharing 法（PSP）の考え方を導入した Q-PSP Learning を提案した．2 つの制御問題を対象にしたシミュレーション実験および移動ロボットを用いた実機での実験を通して，学習の高速化などの点で提案手法の有効性を示した．さらに，連続値入出力を扱うファジィ内挿型に対して Q-PSP Learning の枠組みを導入した手法を提案し，倒立振子制御問題への適用を試みた．提案手法は，環境同定型の強化学習法である Q-Learning に対して，経験強化に徹した Profit Sharing 法の考え方を導入したものであり，学習初期においても報酬を獲得できるため，現実的な問題に対してより有効であると考えられ，環境同定と経験強化のバランスを考慮することが可能な手法といえる．

本章では，強化学習アルゴリズムとして Q-Learning と Profit Sharing 法に着目し，それらを融合した手法を提案したが，本来 Q-Learning はマルコフ性を有する環境に対する手法であり，現実には非マルコフ環境における有効な解法が強く求められている．経験強化に徹した Profit Sharing 法は，そのような環境下でもある程度有効な手法であると考えられ，さらにどのような拡張・発展が必要であるか検討する必要があると思われる．

第 7章

結論

本研究においては，自律的秩序形成および創発的秩序形成の原理に基づくシステムによる「問題解決」に関する考察を行ない，前半では自律的秩序形成に対応する「システム内部における秩序形成による問題解決」のためのシステムを，後半では創発的秩序形成に対応する「システム－環境間における秩序形成による問題解決」のためのシステムを提案した．ここで提案した内容を要約すると以下のようになる．

第 3章では，連続変量やファジィ制約を含む制約充足問題に対して，制約を矩形（直方体）分割することによって制約の分解・還元を行ない，解の構造的選択－上位層での処理－と，解の（連続的な）許容範囲（制約区間）の選択－下位層での処理－に還元することが可能であることを示し，これらの選択を同時並行的かつ重畳した形で実行するシステムとして，全体の論理的整合性を図る記号処理と局所的な相互作用によって解全体のバランスをとる自律分散型処理という二つの計算原理を含む階層型の自律分散システムの枠組を提案した．この枠組による方法は，連続変量・ファジィ制約を含む制約充足問題に対する近似解法となっており，提案システムを並列処理言語 Occam によりトランスピュータ上に実装し，二種類の設計問題への適用を通して，その有効性を明らかにした．

そこで提案した階層型システムは，自律分散システムの構成原理に基づいて，記号処理を行う上位層と連関した形で進行する計算場（下位層）での自己組織化機能に秩序形成（問題解決）を委ねるものであり，自律的秩序形成のためにシナジェティクスの基本原理であるスレイビング原理が発現するように，上位層・下位層間に相互作用を設定した．意志決定者は，システムの実行により得られた解に対して（満足解かどうか）総合的な判断を下すとともに，システム内の種々のパラメータ（初期活性値，リンクの重み，invasion のタイミング等）の設定を行い，意志決定のレベルを一段上げる形で問題解決に関わることになる．もし，満足解でないと判断した場合には，上述のように，意志決定者による介入（システムパラメータの変更）により計算場を再設定した後，問題解決プロセスを再実行することが考えられる．このようなメタレベルにおける意志決定のフィードバックを含んだ問題解決のシス

テム化は，今後の検討課題の一つである．

第 4章では，前章で提案したように，ネットワーク内の全てのリンク構造を同時並行的に活用し，自己組織的・自律分散的に問題解決を行うのではなく，一度に一つのリンク構造だけを選びそれらを逐次的にたどる制約伝播による問題解決プロセスについて検討を行なった．このとき，伝播されるのは区間（制約区間）であり，制約伝播の際に連続量の区間を記号化せずにそのまま伝播する制約伝播力学系においては，安定平衡点に収束する極めて安定的な振る舞いが得られることを数学的に示すとともに，計算機シミュレーションにより確認を行なった．また，制約伝播の際に連続量を記号化（符号化）する手段としてファジィネスを導入したファジィ記号力学系においては，カオス的現象を生み出すような複雑性が内包されていることを計算機シミュレーションを通して明らかにした．

このような複雑な現象が生み出される要因としては，制約伝播力学系の安定的な振舞いに，ファジィラベルの選択という働きかけ（ゆらぎ）が付加されたことが考えられる．つまり，制約伝播の際に記号化のためにファジィラベルの選択のプロセスが介入しているからである．このことは，記号力学の立場から，選択されるファジィラベルの記号列を調べることによっても明らかにした．一般に，記号化を通して力学系などの連続量の状態遷移を構造的に把握する分野は記号力学と呼ばれるが，そこではファジィネスの介入した記号力学系の複雑性の一端を示したといえる．また，このようなファジィ記号力学系によるアプローチと前章で述べた自律分散システムによるアプローチそれぞれの長所を活かし，両者をうまく組み合わせたハイブリッドな問題解決を考えることも可能である．例えば，上位層（記号処理層）からの働きかけを受けつつ下位層（力学層）が中心となって生み出したカオスが，多様な候補解を探索する役割を果たし，その後上位層が適当なタイミングで全体の整合性を見ながら秩序を形成することによって，解を求めることなどが考えられる．具体的には，遺伝的アルゴリズムを用いた上位層での制約処理との融合・併用の可能性などが考えられる．

第 5章では，強化学習の代表的な手法の一つである Q-Learning では従来困難であった連続値の入力（状態）および出力（行動）を扱うことができるように，行動価値関数値である Q 値の導出にファジィ推論を導入した新しい学習法であるファジィ内挿型 Q-Learning を提案した．これはファジィルールを用いて行動価値関数（Q 関数）を表現するものであり，Q 関数をなめらかに近似（内挿）することができ，汎化能力が期待できる．また，最急降下法の導入によりファジィルールのパラメータ・チューニングは前件部・後件部ともに学習プログラムが自律的に行うことができる．さらに，行動が離散値の場合には学習の効率化を図る手法として，各行動ごとにファジィ推論システムを用意するアーキテクチャの導入を提案した．なお，本章では示せなかったが，提案手法では人間がもつ先験的な知識（領域知識）を利用することも容易であり一層の学習の効率化が期待できる．最後に，提案したファジィ内挿型 Q-Learning を倒立振子制御問題と大型船操舵問題の二種類の制御問題に適用し，従来手法と比較することにより，特に学習速度の点においてその有効性を確認した．

そこでは，強化学習アルゴリズムとして Q-Learning に着目し，その拡張をいくつかの面から試みたが，本来 Q-Learning はマルコフ性を有する環境に対する手法であり，現実の多くの問題は非マルコフ的であるため，そのような非マルコフ環境における有効な解法を見い出すことが強く望まれる．1 つの有望な手法としては確率的傾斜法などの記憶を用いないアプローチがあり，それに対してファジィ推論を導入することも考えられる．また，別の手法としてはエージェントの経験した履歴を活用する記憶を用いたアプローチがあり，行動パターンのチャンキングなどによる履歴情報（エピソード）の効率的な活用が考えられる．

第 6章では，前章でも取り上げた Q-Learning に対して，学習の高速化・効率化を図るために，分類子システムで用いられる経験強化型の強化学習アルゴリズムである Profit Sharing 法（PSP）の考え方を導入した Q-PSP Learning を提案した．2 つの制御問題を対象にしたシミュレーション実験および移動ロボットを用いた実機での実験を通して，学習の高速化などの点で提案手法の有効性を示した．さらに，連続値入出力を扱うファジィ内挿型に対して Q-PSP Learning の枠組みを導入した手法を提案し，倒立振子制御問題への適用を試みた．提案手法は，環境同定型の強化学習法である Q-Learning に対して，経験強化に徹した Profit Sharing 法の考え方を導入したものであり，学習初期においても報酬を獲得できるため，現実的な問題に対してより有効であると考えられ，環境同定と経験強化のバランスを考慮することが可能な手法といえる．

そこでは，強化学習アルゴリズムとして Q-Learning と Profit Sharing 法に着目し，それらを融合した手法を提案したが，本来 Q-Learning はマルコフ性を有する環境に対する手法であり，現実には非マルコフ環境における有効な解法が強く求められている．経験強化に徹した Profit Sharing 法は，そのような環境下でもある程度有効な手法であると考えられ，さらにどのような拡張・発展が必要であるか検討する必要がある．

参考文献

[1] 岩井，片井，樋木，坂口，福森：知識システム工学, 計測自動制御学会, 1991

[2] J. R. Anderson: *Cognitive Psychology and Its Implications*, W. H. Freeman and Company（富田訳：認知心理学概論，誠信書房, 1982）

[3] 竹垣盛一，石岡卓也：知的制御システム，海文堂, 1990

[4] P. B. Checkland: *Systems Thinking, Systems Practice*, John Wiley & Sons, Ltd., 1981（高原，中野監訳：新しいシステムアプローチ，オーム社, 1985）

[5] 高原康彦：システム論の基礎，日刊工業新聞社, 1991

[6] 橋田浩一，松原 仁：知能の設計原理に関する試論－部分性・制約・フレーム問題－，日本認知科学会編：認知科学の発展 Vol.7, 共立出版，1993

[7] N. Wiener: *Cybernetics, 2nd edition*, M.I.T. Press, 1961（池原ほか訳：サイバネティクス 第 2 版，岩波書店, 1962）

[8] L. von Bertalanffy: *General System Therory*, George Braziller, 1968（長野，太田訳：一般システム理論，みすず書房, 1973）

[9] I. Prigogine and I. Stengers: *Order out of Chaos*, Bantam Books, 1984（伏見康治ほか訳：混沌からの秩序，みすず書房, 1987）

[10] H. Haken: *Synergetics - An Introduction, Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology*, Springer-Verlag, 1978（牧島，小森訳：協同現象の数理－物理，生物，化学的系における自律形成－，東海大学出版会, 1980）

[11] M. Polanyi: *The Tacit Dimension*, Routledge & Kegan Paul Ltd., 1966（佐藤敬三訳：暗黙知の次元，紀伊国屋書店, 1980）

[12] 星野 力編著：人工生命の夢と悩み，裳華房, 1994

[13] 柴田，福田編：人工生命の近未来，時事通信社, 1994

[14] 北村新三編：特集「創発システム」，計測と制御，Vol.35, No.7, 1996

[15] 伊藤正美編：特集「自律分散システム」，計測と制御，Vol.29, No.10, 1990

[16] 伊藤，市川，須田共編：自律分散宣言－明日を拓くシステムパラダイム－，オーム社，1995

[17] H. A. Simon: *The Science of the Artificial*, M.I.T. Press, 1992（稲葉，吉原訳：システムの科学，パーソナルメディア, 1987）

[18] A. Koestler: *JANUS - A Summing up by Arthur Koestler*, Hutchinson & Co. Ltd., London, 1978（田中，吉岡訳：ホロン革命，工作社, 1983）

[19] 伊藤宏司：生物的工学システム，創発システム公開シンポジウム予稿集，pp.59-60, 1997

[20] 浅間，藤井：ロボットの創発性に関する一考察，第 9 回自律分散システムシンポジウム予稿集，pp.111-114, 1997

[21] 佐倉 統：創発概念の科学史科学哲学的研究，重点領域研究「創発システム」A01 班研究成果報告書「人工生命システム」, 1998

[22] 岩井，片井:問題解決と階層型自律分散システム，計測と制御, Vol.32, No.10, pp.837-842, 1993

[23] 片井 修：システム知能化の考え方の系譜，第 24 回知能システムシンポジウム予稿集, pp.83-88, 1997

[24] 片井 修：新しいファジィ理論が導く関係性の世界とシステム論，日本ファジィ学会誌, Vol.9, No.6, pp.838-850, 1997

[25] D. E. Goldberg: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, 1989

[26] 竹内 勝: 遺伝的アルゴリズムによる機械学習, 計測と制御, Vol.32, No.1, pp.24-30, 1993

[27] 合原一幸編：複雑系が開く世界－科学・技術・社会へのインパクト，日経サイエンス, 1997

[28] 合原一幸：カオス－カオス理論の基礎と応用，サイエンス社，1988

[29] 津田一郎：カオス的脳観，サイエンス社，1990

[30] 金子邦彦：多様性を生み出すカオス，日経サイエンス, No.5, 1994

[31] R. Brooks：A Robust Layered Control System for a Mobile Robot, *IEEE Journal of Robotics and Automation*, Vol.2, pp.14-23, 1986

[32] 谷 淳:力学系に基づく自律移動ロボットの行動学習，日本ロボット学会誌, No.1, Vol.13, 1995

[33] 石田 亨：分散人工知能（1），人工知能学会誌，Vol.7, No.6, 1992

[34] 西原清一：整合ラベリング問題とその応用, 情報処理, Vol.31, No.4, pp.500-507, 1990

[35] 西原清一：制約充足問題の基礎と展望, 人工知能学会誌，Vol.12, No.3, 1997

[36] S. Minton, et al.: Solving Large-Scale Constraint Satisfaction and Scheduling Problems Using a Heuristic Repair Method, *Proc. of the 8th National Conference on Artificial Intelligence (AAAI-90)*, pp.17-24, 1990

[37] 松本, 内野, 狩野, 西原：遺伝的アルゴリズムによる制約充足問題の解法, 情報処理学会研究報告 (AI-101), pp.33-40, 1995

[38] 堀内，半田，渡部，深沢，片井，椹木：有用スキーマを考慮した遺伝的アルゴリズムによる制約充足問題の解法, 第 36 回計測自動制御学会学術講演会予稿集，Vol.2, pp.687-688, 1997

[39] 片井，井田，椹木，岩井：順序制約構造に基づくファジィ集合概念と推論の定式化，計測自動制御学会論文集, Vol.28, No.1, pp.30-39, 1992

[40] 片井，井田，椹木，岩井：区間制約ファジィ推論による倒立振子の制御と車両移動計画の生成，計測自動制御学会論文集，Vol.29, No.4, pp.470-479, 1993

[41] 増市，片井，椹木，西山，岩井：問題解決のための階層型自律分散システムの構成，計測自動制御学会論文集, Vol.28, No.11, pp.1364-1373, 1992

[42] 片井，松原，井田，椹木，増市，岩井：ファジィ変量を含む制約充足問題に対する分散型問題解決の枠組み，第 15 回知能システムシンポジウム予稿集, pp.75-80, 1992

[43] 片井，松原，増市，椹木，片山，岩井：記号処理・ファジィ推論融合型問題解決の枠組み，第9回ファジィシステムシンポジウム講演論文集，pp.521-524, 1993

[44] D. Pountain and D.May: *A Tutorial Introduction to Occam Programming*, INMOS Limited, 1987

[45] R. Seidel: A New Method for Solving Constraint Satisfaction Problems, *Proc. of IJCAI'81*, pp.338-342, 1981

[46] O. Katai, et al.: An Autonomous Decentralized System for Constraint-Oriented Problem Solving Involving Continous and Fazzy Variables, *Proc. of the 1st International Symposium on Autonomous Decentralized Systems (ISADS'93)*, pp.63-69, 1993

[47] B. Freeman-Benson et al.: An Incremental Constraint Solver, *Communications of the ACM*, Vol.33, No.1, pp.54-62, 1990

[48] E. Freuder: Partial Constraint Satisfaction, *Proc. of IJCAI'89*, pp.278-283, 1989

[49] O. Katai, M. Ida, T. Sawaragi and S. Iwai: Dynamic and Context-Dependent Treatment of Fuzziness from Constraint-Oriented Perspectives, *Proc. of the 4th International Fuzzy Systems Association World Congress (IFSA'91)*, Vol. on Artificial Intelligence, 1991

[50] 片井，堀内，椹木，岩井，平岡：ファジィ記号力学系のカオス構造と問題解決，第10回ファジィシステムシンポジウム講演論文集，pp.97-103, 1994

[51] E. Tsang: *Foundations of Constraint Satisfaction*, Academic Press, 1993

[52] D. R. Smart: Fixed point theorems, *Cambridge Tracts in Mathematics*, Vol.66, Cambridge Univ. Press, 1977

[53] H. Bai-lin: *Elementary Symbolic Dynamics*, World Scientific, 1989

[54] R. Devaney: *An Introduction to Chaotic Dynamical Systems*, Addison-Welsley, 1989

[55] P.Berge, Y.Pomeau, Ch.Vidal；*L'ordre dans le Chaos : Vers une Approche Deterministe de la Turbulence*, 1984（相澤洋二訳：カオスの中の秩序，産業図書，1992）

[56] Samuel, A. L.: Some Studies in Machine Learning Using Game of Checkers, *IBM Journal of Research and Development*, Vol.3, pp.210-229, 1959

[57] Kaelbling, L. P.: *Learning in Embeded Systems*, M.I.T. Press, 1993

[58] Sutton, R. S.: Learning to Predict by the Methods of Temporal Difference, *Machine Learning*, Vol.3, pp.9-44, 1988

[59] Watkins, C. J. and Dayan, P.: Technical Note: Q-Learning, *Machine Learning*, Vol.8, pp.279-292, 1992

[60] Peng, J. and Williams, R. J.: Incremental Multi-Step Q-Learning, *Proc. of the 11th International Conference on Machine Learning*, pp.226-232, 1994

[61] Holland J. H. et al.: *INDUCTION*, M.I.T. Press, 1986

[62] Grefenstette, J. J.: Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms, *Machine Learning*, Vol.3, pp.225-245, 1988

[63] Saito, F. and Fukuda, T.: Two-Link-Robot Brachiation with Connectionist Q-Learning, *From Animals to Animats 3*, The M.I.T. Press, pp.309-314, 1994

[64] Lin, Long-Ji: Self-Improving Reactive Agents: Case Studied of Reinforcement Learning Frameworks, *From Animals to Animats*, The M.I.T. Press, pp.297-305, 1992

[65] 菅野道夫：ファジイ制御，日刊工業新聞社, 1988

[66] Cashwell, E. D. and Everett, C. j.: *Monte-Carlo Method*, Pergamon Press, 1959

[67] Rumelhart, D. E. and MaClelland, J. L. (eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Congnition*, M.I.T. Press, 1988

[68] 野村，林，若見：デルタルールによるファジィ推論の自動チューニング手法と障害物回避への応用，日本ファジィ学会誌，Vol.4, No.2, pp.379-388, 1992

[69] Jouffe, L.: Ventilation Control Learning with FACL, *Proc. of the 6th IEEE International Conference on Fuzzy Systems*, Vol.3, pp.1719-1724, 1997

[70] Barto, A.G., et.al.: Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Ploblems, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.13, No.5, pp.834-846, 1983

[71] 小池，畝見，古谷：遺伝的アルゴリズムを用いた遅れ系の制御，第 15 回知能システムシンポジウム予稿集，pp.7-12, 1992

[72] 木村，宮崎，小林：確率的傾斜法による強化学習：不完全知覚問題への接近，システム／情報合同シンポジウム'96 予稿集, pp.63-68, 1996

[73] Motoda, H. et.al.: Extracting Behavioral Patterns from Relational History Data, *Proc. of the 6th International Conference on User Modeling*, 1997

[74] K-Team SA: Khepera User Manual

[75] McCallum, R. A.: Hidden State and Reinforcement Learning with Instance-Based State Identification, *IEEE Trans. of Systems, Man, and Cybernetics - Part B: Cybernetics*, Vol.26, No.3, pp.464-473, 1996

[76] T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi: Fuzzy Interpolation-Based Q-Learning with Continuous States and Actions, *Proc. of the 5th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'96)*, Vol.1, pp.594-600, 1996

[77] T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi: Q-PSP Learning: An Exploitation-Oriented Q-Learning Algorithm and Its Applications, *Proc. of the 3rd IEEE International Conference on Evolutionary Computation (ICEC'96)*, pp.76-81, 1996

[78] Kaelbling, L. P., et al.: Reinforcement Learning: A Survery, *Journal of Artificial Intelligence Research*, Vol.4, pp.237-277, 1996

[79] 畝見達夫：強化学習, 人工知能学会誌, Vol.9, No.6, pp.831-836, 1994

[80] 山村，宮崎，小林：強化学習の特徴と発展の方向，システム制御情報学会誌，Vol.39, No.4, pp.191-196, 1995

[81] 山村，宮崎，小林：エージェントの学習，人工知能学会誌，Vol.10, No.5, pp.683-689, 1995

[82] 宮崎，小林：離散マルコフ決定過程下での強化学習，人工知能学会誌，Vol.12, No.6, pp.811-821, 1997

[83] 宮崎，山村，小林：強化学習における報酬割当ての理論的考察，人工知能学会誌，Vol.9, No.4, pp.580-587, 1994

著者関連文献

学術論文

1. 片井，堀内，松原，椹木，岩井：連続変量・ファジィ制約を含む制約充足問題解決のための自律分散システム，計測自動制御学会論文集，Vol.31, No.5, pp.640-649, 1995

2. 堀内，藤野，片井，椹木：連続値入出力を扱うファジィ内挿型 Q-Learning の提案，計測自動制御学会論文集（投稿中）

3. 堀内，藤野，片井，椹木：経験強化を考慮した Q-Learning の提案とその応用，計測自動制御学会論文集（投稿中）

国際会議

1. T. Nishiyama, O. Katai, T. Sawaragi, S. Iwai and T. Horiuchi: A Framework for Multiagent Planning and a Method of Representing its Plan Integration, *Proc. of the 4th Transpuer/Occam International Conference*, pp.146-160, 1992

2. O. Katai, T. Nishiyama, T. Sawaragi, T.Horiuchi and S. Iwai: Decentralized Control of Discrete Event Systems Based on Extended Higher Order Petri Nets, *Proc. of the 1st Asian Control Conferences (ASCC'94)*, Vol.2, pp.897-900, 1994

3. O. Katai, T. Horiuchi, S. Matsubara, T. Sawaragi and S. Iwai: Decentralized Constraint-Oriented Problem Solving Based on Fuzzy Coding of Complex Constraints Involving Continuous Variables, *Proc. of the 6th International Fuzzy Systems Association World Congress (IFSA'95)*, Vol.1, pp.133-136, 1995

4. O. Katai, T. Horiuchi, T. Sawaragi, S. Iwai and T. Hiraoka: Chaotic Structure of Fuzzy Symbolic Dynamics and Their Relation to Constraint-oriented Problem Solving, *Proc. of the International Joint Conference of the 4th IEEE International Conference on Fuzzy Systems and the 2nd International Fuzzy Engineering Symposium (FUZZ-IEEE/IFES'95)*, Vol.4, pp.1955-1962, 1995

5. T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi: Fuzzy Interpolation-Based Q-Learning with Continuous States and Actions, *Proc. of the 5th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'96)*, Vol.1, pp.594-600, 1996

6. T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi: Q-PSP Learning: An Exploitation-Oriented Q-Learning Algorithm and Its Applications, *Proc. of the 3rd IEEE International Conference on Evolutionary Computation (ICEC'96)*, pp.76-81, 1996

7. O. Katai, T. Nishiyama, T.Horiuchi and T. Sawaragi: Hypermedia-based Process Operation Support System:, *Proc. of the International Conference on Cognitive Systems Engineering in Process Control (CSEPC'96)*, pp.208-215, 1996

8. T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi: Fuzzy Interpolation-Based Q-Learning with Profit Sharing Plan Scheme, *Proc. of the 6th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'97)*, Vol.3, pp.1707-1712, 1997

9. O. Katai, T. Horiuchi and T. Sawaragi: The Cause And Meaning of Chaos Phenomena in Symbiotic Problem Solving Uniting Natural System Redundancy and Artificial System Logicality, *Proc. of the International Symposium on System Life (ISSL'97)*, pp.253-262, 1997

10. O. Katai, T. Higuchi, T.Horiuchi and T. Sawaragi: Self-Organizing Fuzzy Control Systems by GA-Based Group Selection of Constraint-Interval Control Rules, *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97)*, Vol.3, pp.249-254, 1997

11. T. Shiose, T. Sawaragi, O. Katai and T.Horiuchi: Proactive Behavior Formation of Autonomous Robots with Evolutional Perception, *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97)*, Vol.4, pp.474-479, 1997

書籍（分担執筆）

1. O. Katai, M. Ida, T. Sawaragi, K. Shimamoto, S. Iwai, T. Horiuchi and M. Terabe: Fuzzy Control as Self-Organizing Adaptive Constraint-Oriented Problem Solving, in Z. Bien and C. Min (eds.): *Fuzzy Logic and its Applications to Engineering, Infor-*

mation Science, and Intelligent Systems, pp.111-132, Kluwer Academic Publishers, 1995

解説記事

1. 堀内，片井：A*アルゴリズム，日本ファジィ学会誌，Vo.7，No.6，pp.1149-1154，1995

謝辞

本研究をとりまとめるにあたり，懇切なる御指導，御配慮を賜り，多大な御尽力をいただいた京都大学大学院情報学研究科 片井 修教授に心から感謝申し上げます．

また，有益な御教示と御指示，御鞭撻を賜った京都大学名誉教授 現名城大学理工学部 岩井壮介教授ならびに京都大学大学院工学研究科 椹木哲夫助教授に深甚な謝意を表します．

また，文部省科学研究費重点領域研究「創発システム」等を通して有益なご助言をいただきました京都大学大学院工学研究科 土屋和雄教授ならびに東京工業大学大学院総合理工学研究科 小林重信教授に深く感謝いたします．

また，平素から有益な御教示と御指導をいただいた大阪大学産業科学研究所 元田 浩教授ならびに鷲尾 隆助教授に深く感謝いたします．

さらに，本研究を遂行するにあたり，暖かい御支援と御激励をいただいた井田正明助手をはじめとする京都大学大学院情報学研究科片井研究室の皆様，ならびに京都大学大学院工学研究科椹木研究室の皆様，大阪大学産業科学研究所元田研究室の皆様に心よりお礼申し上げます．

最後に，常に優しく見守るとともに暖かい励ましをいただいた両親に心から感謝いたします．